THE UNIVERSITY
of EDINBURGH

SYSTEMS ANALYSIS OF THE HUMAN CELL CYCLE TRANSCRIPTION NETWORK

Sz-Hau Chen

Thesis submitted for the degree of
Doctor of Philosophy

College of Medicine and Veterinary Medicine,
University of Edinburgh

2016

Abstract

Cell division is one of the most fundamental processes of life whereby one cell replicates itself to produce two. The molecular machinery that drives and regulates this fundamental process has been much studied but much remains unknown. This work describes the use of transcriptomics analyses to identify putative new proteins involved with this process and subsequent attempts to prove their association with this pathway. Using the latest array technology, in Chapter 2 I describe studies that examine the expression of genes regulated during different stages of the human cell cycle. Synchronous populations of neonatal human dermal fibroblasts (NHDFs) were generated by serum starvation and analysed in two separate microarray experiments. For the first set array experiments, samples were taken every 6 hours for 48 hours after serum refeeding, and every 2 hours for 24 hours for the second experiment. Using BioLayout *Express*^{3D}, network structure analyses identified four major clusters of gene expression patterns associated with different stages of the cell cycle: G_0-, early G_1-, late G_1-, and S/G_2/M-phase. By comparison with datasets of other human cells and tissues, the list of genes in the S/G_2/M cluster was refined; genes were only kept in the list if they were found to be co-expressed in cells and tissues with high levels of cell proliferation. 706 genes that were co-expressed during S/G_2/M-phase were selected for further analyses. Manual curation showed that 484 are known cell cycle-associated genes, 78 are genes with putative association to the cell cycle, and 75 have known roles in other biological processes, whilst 69 were entirely uncharacterised genes. In order to investigate the 69 genes with unknown function, in Chapter 3 I describe how RNAi was used to screen 42 of these genes to see if their knockdown resulted in an effect on cell proliferation. After extensive assay optimisation, endoribonuclease-prepared siRNA (esiRNA) was delivered to NHDF cells and the effect of knockdown determined using a real time cell analysis (RTCA) system. This system monitors the change in electrical resistance induced by growing cell populations defined as the cell impedance index (CI). Using a Z-scoring cut-off to determine the hits of the RNAi screening, according to the average value of cell impedance growth rate (CIGR i.e. a value from transformed CI), 19 of 42 genes were found to significantly affect the dynamics of cell proliferation, supporting a potential role in cell division. In order to verify that the unknown proteins localise to structures compatible with a role in the cell cycle, in Chapter 4 I describe protein localisation studies on 11 of 19 genes of 'hits' from Chapter 3 (we were unable to obtain clones for the other 8 genes) and other genes of interest. Transfection studies of HEK293T cells with expression clones containing more than 11 ORFs with GFP fused to either the N- or C-terminal were performed. FAM111B and

KIAA1549L appeared to be localised to the centrosome. In order to better understand the context in which the novel centrosomal proteins that FAM111B might operate, in Chapter 5 I describe the construction of a large-scale pathway model of centrosome life cycle based on an extensive literature review. The model is composed of 117 of the most important centrosome-associated proteins and has been constructed using the modified Edinburgh Notation (mEPN) scheme. This model was used to better annotate the genes in the original S/G_2/M list and understand which of the genes in the model are regulated during cell division. This regulatory network model of the centrosome life cycle represents an important summary of current knowledge and provides a useful resource for further analyses of the novel centrosomal proteins.

In summary, a list cell cycle gene was derived from microarray experiments by using network structure analyses. Subsequent analyses filtered the genes that co-expressed during S/G_2/M-phase narrowing down into 706 genes. Of this list, 69 genes had not previously been associated with the cell cycle. 42 of these unknown genes were analysed by using real time RNAi screening, 19 of these genes were indeed associated with the cell proliferation, and 2 of these genes with unknown function appear to localise to the centrosome. To predict their involvement in the centrosome life cycle, a pathway map composed of 117 centrosome-associated proteins were formed. Although further research is needed to determine their position in the centrosome life cycle, the pathway can be used for computational modelling testing their putative function in the system.

Lay summary

All living things (e.g. plants and animals) are composed of cells. For an organism to grow and repair itself, cells must divide to produce two daughter cells, a process called cell division or mitosis. Cell division can itself be divided up into a number of distinct stages. First the cell must grow in size. This is a stage called gap-phase 1 (G_1-phase). It must then copy its DNA (S-phase). There is a second gap phase (G_2-phase) during which the cells prepare for cell division and check that all is in order before their physical division (M-phase). Depending on conditions, cells may stop from dividing further (when they are said to enter G_0-phase) or if conditions are favourable re-enter the cell cycle and repeat the process. To properly regulate the complicated biological machinery behind the cell cycle, a large number of proteins are involved. In the past 20 years, many of their roles in the cell cycle have been studied. However, some of them remain unknown. In the past, largely due to limitations of technology, scientists could only study particular biological events controlled by proteins and genes individually. Recent advancements in technology, particularly the development of the microarray, have revolutionized the study of cell regulatory mechanisms. Microarrays allow scientists to monitor thousands of biological events by detecting the activity (expression) of all genes at the same time. By using the latest technology, I performed two microarray experiments following the gene expression of human cells undergoing cell division. Four major groups of genes were identified whose activity was associated with a specific stage of the cell cycle. Of particular interest were the genes whose expression was associated with S-, G_2-, and M-phase. By studying the expression of these genes in many other types of cells and tissues, I have focused on 706 genes to be of particular interest. By searching the published literature and online resources, many of these genes were found to be the known cell cycle genes, however 69 of them were previously uncharacterised genes. Therefore I decided to investigate the biological function of genes in more detail. To further investigate their role in cell proliferation, I knocked down the activity of a number of the uncharacterised genes to see the effect on cell proliferation using a technique called RNAi. Out of 42 genes tested, 19 were found to alter the rate of proliferation adding support to the idea that they may play a yet to be defined role in cell division. In order to characterise their biological function further, I localised proteins encoded by 11 of 19 uncharacterised genes and other genes of interest by tagging them with green fluorescence proteins (GFP). The location of the florescence signal was used to find the position of the proteins in the cells using a microscope. FAM111B and KIAA1549L appeared to be localised to the centrosome. This implied the possibility that the unknown

proteins of interest may potentially be involved in the centrosome life cycle. In order to better understand the role of the centrosome and therefore the novel proteins associated with it, I have built a pathway map of the centrosome life cycle. This pathway map of the centrosome life cycle includes the 117 most important centrosomal proteins. However further experimental approaches are required to position these two unknown proteins owing to poor information from the published literature. In summary, I have reanalysed the gene expression events associated with the cell cycle, tested a number of the novel genes associated with the cell proliferation by RNAi screening. Two proteins with unknown function appeared to be involved in centrosomal activity according to their subcellular localisation and important in cell proliferation when their activities were turn-off. A large-scale pathway map was developed in order to help understand their position in the centrosome life cycle.

Declaration

The thesis presented is the work of the author except where stated otherwise by reference and/or acknowledgement. Any work presented, which has been conducted by (or in collaboration with) others is explicitly acknowledged at the end of each Chapter. No part of this work has been submitted in candidature for any other degree or qualification.

Name: Sz-Hau Chen  Date: 10/April/2016

Contents

Chapter 1.

Chapter 3.

RNA interference targeting of novel cell cycle-associated genes

Chapter 5.

Construction of a large-scale molecular model of centrosome and its duplication during cell division

Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Tables

Abbreviations

AAAS	achalasia, adrenocortical insufficiency, alacrimia
AKAP9	a kinase anchor protein 9
ANAPC	anaphase promoting complex
ASPM	asp (abnormal spindle) homolog, microcephaly associated
ATAD2	ATPase family, AAA Domain containing 2
ATP	adenosine triphosphate
AURKA	aurora kinase A
BBS4	bardet-biedl syndrome 4
BUB1	budding uninhibited by benzimidazoles 1
C17orf53	chromosome 17 open reading frame 53
CaMKII	Ca^{2+}/ calmodulin-dependent protein kinase II
cAMP	cyclic adenosine monophosphate
CC2D1A	coiled-coil and C2 domain containing 1A
CCNA2	cyclin A2
CCNB1	cyclin B1
CCNB2	cyclin B2
CCND1	cyclin D1
CCNF	cyclin F
CCP110	centriolar coiled-coil protein of 110 kDa
CDC20	cell division cycle 20
CDC25A	cell division cycle 25A
CDC25B	cell division cycle 25B
CDC6	cell division cycle 6
CDC7	cell division cycle 7
CDH1	cadherin 1
CDK	cyclin-dependent kinase
CDK1	cyclin-dependent kinase 1
CDK2	cyclin-dependent kinase 2
CDK4	cyclin-dependent kinase 4
CDK5RAP2	CDK5 regulatory subunit associated protein 2
CDK6	cyclin-dependent kinase 6
CDK7	cyclin dependent kinase 7
CDKN1A	cyclin-dependent kinase inhibitor 1A
CDKN2B	cyclin-dependent kinase inhibitor 2B
cDNA	complementary DNA
CDT1	chromatin licensing and DNA replication factor 1
CENPF	centromere protein F
CENPJ	centromere protein J
CEP120	centrosomal protein 120kDa
CEP135	centrosomal protein 135kDa
CEP152	centrosomal protein 152kDa

CEP164	centrosomal protein 164kDa
CEP170	centrosomal protein 170kDa
CEP192	centrosomal protein 192kDa
CEP290	centrosomal protein 290kDa
CEP350	centrosomal protein 350kDa
CEP55	centrosomal protein 55kDa
CEP57	centrosomal protein 57kDa
CEP63	centrosomal protein 63kDa
CEP72	centrosomal protein 72kDa
CETN1	centrin-1
CI	cell impedance index
CIGR	cell impedance index growth rate
CLSPN	claspin
CNN	centrosomin
CNTRL	centriolin
CROCC	ciliary rootlet coiled-coil rootletin
CTCF	CCCTC-binding factor
DAVID	database for annotation, visualization and integrated discovery
DBF4	DBF4 Zinc Finger
DCTN1	dynactin 1
DERL2	Derlin 2
DISC1	disrupted in schizophrenia 1
DNA	deoxyribonucleic acid
DNAH2	dynein axonemal heavy chain 2
dsRNA	double-stranded RNA
E2F4	E2F transcription factor 4
E2F5	E2F transcription factor 5
ER	endoplasmic reticulum
esiRNAs	endoribonuclease-prepared siRNAs
EtOH	ethanol
FA	fanconi anemia
FANCA	fanconi anemia, complementation group A
FBS	fetal bovine serum
FBXW5	f-box and WD repeat domain containing 5
FNTB	farnesyltransferase, CAAX box, beta
FOPNL	FGFR1OP N-terminal like
FZR1	fizzy/cell division cycle 20 related 1
GFP	green fluorescence protein
GO	gene ontology
HAUS1	HAUS augmin-like complex, subunit 1
HEK293T	human embryonic kidney 293T
HGNC	human gene nomenclature
HIRA	histone cell cycle regulator

HIST1H1A	histone cluster 1, H1a
HOOK3	hook microtubule-tethering protein 3
HTS	high throughput system
HUGO	human genome organisation
IFM	immune fluorescent microscopy
IL-2	interleukin 2
ING4	inhibitor of growth family, member 4
INK4	inhibitors of CDK4
KATNB1	katanin p80 subunit B1
KEGG	kyoto encyclopaedia of genes and genomes
KIF11	kinesin family member 11
KIF15	kinesin family member 15
KIF24	kinesin family member 24
KIZ	kizuna centrosomal protein
KPNB1	karyopherin β1
LTBP	latent transforming growth factor
LTBP2	latent transforming growth factor beta binding protein 2
MAD2L1	MAD2 mitotic arrest deficient-like 1
MAPRE1	microtubule-associated protein, RP/EB family member 1
MCL	Markov CLustering
MCM	minichromosome maintenance
MCM6	minichromosome maintenance complex component 6
mEPN	modified Edinburgh Pathway Notation
MNAT1	MNAT CDK-activating kinase assembly factor 1
MT	microtubule
MTOC	micro tubule-organising centre
MYT1	myelin transcription factor 1
MZT1	mitotic spindle organizing protein 1
MZT2	mitotic spindle organizing protein 2
NEBD	nuclear envelope break down
NEDD1	neural precursor cell expressed, developmentally down-regulated 1
NHDF	neonatal human dermal fibroblast
NIN	ninein
NPM1	nucleophosmin family member 1
NUMA1	nuclear mitotic apparatus protein 1
ODF1	outer dense fibre of sperm tails 1
ODF2	outer dense fibre of sperm tails 2
OFD1	oral-facial-digital syndrome 1
ORC	origin recognition complex
ORC5	origin recognition complex, subunit 5
ORFs	open reading frames
PAFAH1B1	platelet-activating factor acetylhydrolase 1B regulatory subunit 1
PARD6A	partitioning defective 6 α

PCM	pericentriolar material
PCM1	protein pericentriolar material 1
PCNA	proliferating cell nuclear antigen
PCNT	pericentrin
PLK	polo-like kinase
PLK2	polo-like kinase 2
PLK4	polo-like kinase 4
POC1A	proteome of centriolar protein A
POC1B	proteome of centriolar protein B
POC5	proteome of centriolar protein 5
POLA1	polymerase (DNA directed), α1
PP2A	protein phosphatase 2 α
PPP1CA	protein phosphatase 1 catalytic subunit α
PPP1CC	protein phosphatase 1 catalytic subunit
PPP2R1B	protein phosphatase 2, regulatory subunit A, beta
PRC1	Protein regulator of cytokinesis 1
pre-RCs	pre-replicative complexes
PTGS	post-transcriptional gene silencing
RAD21	double-strand-break repair protein rad21 homolog
RAD51C	RAD51 paralog C
RBM22	RNA binding motif protein 22
RISC	RNA induced silencing complex
RMA	robust multi-array average
RNAi	RNA interference
RRM2	protein ribonucleotide reductase M2
RTCA	real time cell analysis
SASS6	spindle assembly 6
SAV1	salvador family WW domain containing protein 1
SBGN	systems biology graphical notation
SGOL1	cohesin-associated protein shugoshin-like 1
shRNAs	short-hairpin RNAs
siRNA	small interfering RNA
SKA1	spindle and kinetochore associated complex subunit 1
SPAG5	sperm associated antigen 5
ssRNAs	single-stranded RNAs
STAT5A	Signal transducer and activator of transcription 5A
STIL	SCL/TAL1 interrupting locus
STK3	serine/threonine kinase 3
SV40	simian vacuolating virus 40
TEKT1	tektin 1
THP-1	human acute monocytic leukemia cell line
TM7SF2	transmembrane 7 superfamily member 2
TOP2A	topoisomerase (DNA) II α

TP53	tumour protein p53
TPX2	TPX2, microtubule-associated
TSG101	tumour susceptibility 101
TUBE1	tubulin, epsilon 1
TUBG1	tubulin gamma 1
TUBG2	tubulin gamma 2
TUBGCP2	tubulin gamma complex associated protein 2
TUBGCP3	tubulin gamma complex associated protein 3
USP33	ubiquitin specific peptidase 33
UTR	untranslated region
XRCC1	X-ray repair complementing defective repair in Chinese hamster cells 1
γ-TuRC	γ-tubulin ring complex
γ-TuSC	γ-tubulin small complex

Dedication

This thesis is dedicated to my parents with unconditional love and support at any levels.

Acknowledgements

The analyses described in this thesis were performed at The Roslin Institute (University of Edinburgh). I very much appreciate the help of my supervisor Professor Tom Freeman and co-supervisor Dr. Mark Barnett. They helped me through countless challenges, with their patience, enthusiasm, wonderful guidance of my research, and of course for reading and re-reading drafts of my thesis. Tom also gave me the opportunity to attend the workshop at the Wellcome Trust Advanced Course demonstrating the power of BioLayout *Express*^{3D} and array data mining. This provided priceless experience enhancing my social techniques when interacting with participants from multiple cultures. I'm particularly grateful to Dr. Mark Barnett for his training in laboratory techniques (i.e. watch one, practice one and teach one) and his great patience in helping me with setting up large and complex experiments, and for academic writing training as well. I also thank Dr. Emily Clark and Freeman group for assistance of my thesis corrections at many levels as well.

Personal

I am indebted to my parents for their constant support, encouragement, and endless patience. Thanks also to my baby sister Vivian Chen. Edinburgh has been a great city to conduct an academic career for a PhD and I have made friends of different nationalities from around the world.

Publication from previous work

Study on diversity of rice sheath blight in Taiwan

Szu(Sz)-Hao(Hau) Chen, Hsiang-Chen Yang, Tse-Chun Fang, Chien-Hua Chang, I.J. Lin, and Lung-Chung Chen (The 9th bilateral meeting of Taiwan and China on fungi research Page 207–226)

Manuscripts in preparation

Assembly of a Parts List of the Human Cell Cycle

Sz-Hau Chen*, Bruno Giotti*, Mark Barnett*, Fahmi wan Nazarie, David A. Hume, Stefan Weimann and Tom C. Freeman† [* equal first author]

Construction of a large-scale molecular model of the centrosome and its replication during cell division

Sz-Hau Chen, Mark Barnett, and Tom C. Freeman

A graphical and computational system for modelling the dynamics of biological pathways and other complex systems

Laura O'Hara, Alessandra Livigni, Thanasis Theocharidis, Benjamin Boyer, Tim Angus, Derek Wright, Sz-Hau Chen, Sobia Raza, Mark W. Barnett, Paul Digard, Lee B. Smith, and Tom C. Freeman

Graph-based visualisation and analysis of RNA-seq data

Fahmi W. Nazarie, Tim Angus, Sz-Hau Chen, Mark Barnett, Karsten Klein, Anton J. Enright, Tom C. Freeman

Abstracts

Posters

Poster presentations at "Functional Genomics and Systems Biology" conference: Identification and characterisation of novel cell cycle genes

Sz-Hau Chen, Mark Barnett, Bruno Giotti, Fahmi wan Nazarie, Tom C. Freeman 2013.
The Wellcome Trust Genome Campus, Hinxton, Cambridge

Analysis and modelling of the core protein components of the cell cycle

Sz-Hau Chen, Mark Barnett, Tom C. Freeman 2012. Student's day

Identification and characterisation of novel cell cycle genes

Sz-Hau Chen, Mark Barnett, Bruno Giotti, Tom C. Freeman 2013. Student's day

Oral presentation

Identification and characterisation of novel cell cycle genes by RNAi

Sz-Hau Chen, Mark Barnett, Bruno Giotti, Fahmi wan Nazarie, Tom C. Freeman 2014. Student's day

Chapter 1.

The cell cycle

Introduction to the cell cycle

Only by duplication of parent cells is life able to continue. Cells are considered as the basic units of life. The concept of the cell playing a role as the basic element of the living organism was first suggested and officially formulated during the 1830s. This eventually became known as cell theory and included three basic components. Firstly, all organisms are composed of one or more cells [1-3]. Secondly, in all living creatures, the cell is the basic unit [1-3]. Thirdly, all types of cells are produced by the proliferation and division of parental cells [3,4].

Cell theory was known to biologists for over 100 years before the pioneering work of Alma Howard and Stephen Pelc suggested that deoxyribonucleic acid (DNA) replication was only activated at a particular phase of the cell cycle and did not occur during mitosis [5]. Their work in the broad bean (*Vicia faba*) suggested that the cell cycle was composed of a number of stages before and after cell division [6]. Based on this discovery, cell biologists were able to define five different stages of the cell cycle based on their features at the molecular level, including G_0-phase (gap with 0 growth), G_1-phase (gap 1 phase), S-phase (DNA synthesis), G_2-phase (gap 2 phase), and M-phase (mitosis) (Figure 1.1). The cycle of cell duplication and division is therefore termed the cell cycle and is the essential mechanism by which all living creatures reproduce.

Eukaryotic cells have a complicated regulatory network of proteins controlling the cell cycle. The aim of this regulation is to accurately duplicate the DNA in chromosomes and then conduct the segregation of sister chromatids into two daughter cells (Figure 1.1). The cell cycle is composed of two major phases, interphase and M-phase (Figure 1.1). Interphase is composed of G_1-, S-, and G_2-phase (Figure 1.1). M-phase can be divided into prophase, prometaphase, metaphase, anaphase, telophase, and cytokinesis (Figure 1.1). During cytokinesis, the cytoplasm of a parental cell is physically divided into the two daughter cells. For typical mammalian cells, the interphase might require 23 h of a 24 h long cell cycle with M-phase taking one hour to complete.

In preparation for S-phase, most cells require time to grow and increase their amounts of cellular constituents (e.g. proteins and organelles) before DNA duplication and cell division.

This is called the G_1-phase. The duration of G_1-phase is between 8 and 16 h for most rapidly dividing cell lines (doubling time < 30 h) [7] e.g. the doubling time of fibroblast is approximately 25 h [8]. In contrast, some transformed cells e.g. acute monocytic leukemia cell line (THP-1) have longer doubling times between 60 and 70 h [9]. For DNA duplication, S-phase needs less than 8 to 12 h, G_2-phase requires less than 2 to 4 h, and less than one hour is needed for cytokinesis [7,10].

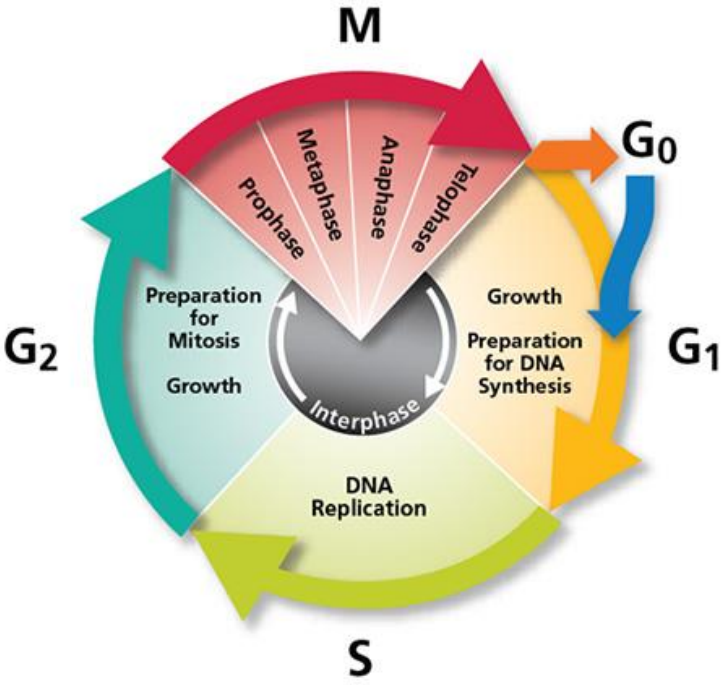


Figure 1.1: The eukaryotic cell cycle is divided into a number of stages. Two gap phases (G_1 and G_2) allow cell growth in preparation for DNA duplication and chromosome segregation during interphase. DNA duplication is triggered after G_1-phase and finished by the end of the S/G_2 border, M-phase is composed of four stages, Prophase, Metaphase, Anaphase, and Telophase. This figure is courtesy of BD Biosciences Inc. [11].

Cell cycle control system

The system controlling the cell cycle responds to the internal and external environment of the cell determining the progression or delay of the cell cycle (Figure 1.2). Progression through each phase of the cell cycle is monitored by checkpoints that can detect problems with cell growth, DNA replication, and mitotic sister chromatids segregation.

These checkpoints are comprised of signalling cascades, which can block the progression of the cell cycle until recovery has been achieved. The first checkpoint, called the 'start checkpoint', is positioned at late G_1-phase, where the conditions, i.e. cell size, nutrients supplied from the environment and DNA integrity, determine whether the cell commits to enter the cell cycle (Figure 1.2) [12].

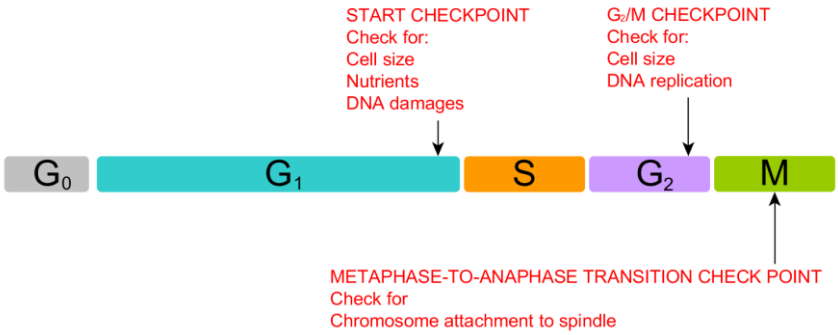


Figure 1.2: Cell cycle control and the checkpoint system. This diagram depicts the checkpoint system that can detect irregular biological events and block the progression of the cell cycle. This figure was adapted from 'The Cell Cycle, principle of control' by Dr. D. O. Morgan [13].

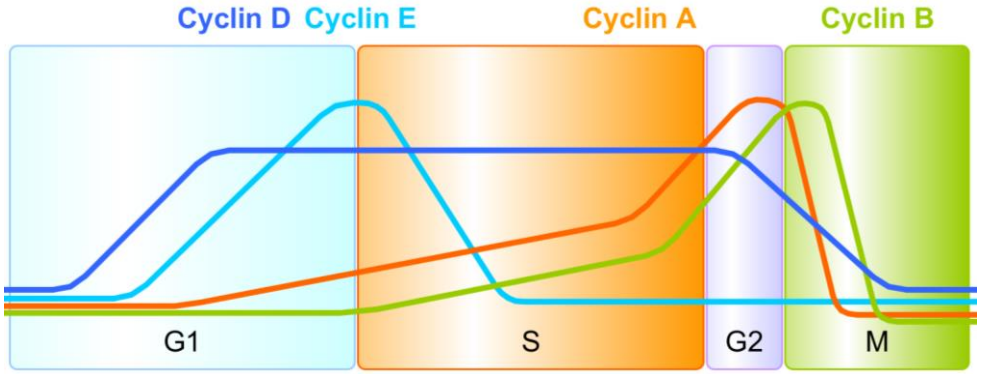
The G_2/M-checkpoint is the second checkpoint that controls initiation of events at an early stage of M-phase that eventually lead to alignment of duplicated chromosomes at metaphase (Figure 1.2). The size of cells and DNA duplication determine whether the cell commits to enter M-phase. The third checkpoint is located at the transition between metaphase and anaphase, and monitors the segregation of sister chromatids. The attachment between the mitotic chromosomes and spindles determines whether the cell commits enter the cytokinesis (Figure 1.2). Inappropriate environments that restrict cell growth result in blocking the cell cycle at the checkpoint near G_1/S border [12]. Incomplete DNA duplication blocks the cell cycle at G_2/M-checkpoint [14,15]. During the last checkpoint (i.e. metaphase to anaphase transition), the duplicated sister chromatids are segregated into two daughter cells, committing a cell to be divided into two daughter cells (Figure 1.2) [16,17].

Cyclins and CDKs

Little was known about the complexity of the cell cycle regulatory network until the discovery of cyclins in the 1980s [18,19]. Mitotic cyclins were first identified in sea urchins, the synthesis and periodic degradation of these proteins were detected shifting between interphase and M-phase [20]. These proteins were given their name because of their periodical synthesis and degradation through different phases of the cell cycle [20].

Cell cycle regulation is driven by a family of cyclins and their corresponding cyclin-dependent kinases (CDKs). It is known that various combinations of cyclins and CDKs trigger the progression of the cell cycle through its different phases [21]. In contrast to the cyclins, which are periodically synthesized and degraded, the level of CDK proteins remains constant across the cell cycle.

A



B

Cyclin/CDK Complex	Cyclin	CDK Partner	
G1-CDK	Cyclin D	CDK4	CDK6
G1/S-CDK	Cyclin E	CDK2	
S-CDK	Cyclin A	CDK2	CDK1
M-CDK	Cyclin B	CDK1	

Figure 1.3: Periodical regulation of Cyclin/CDK complexes in the cell cycle. A) Activity of the cyclins in different phases of the cell cycle. **B)** Cyclins and their partner CDKs during different stages of the cell cycle. This figure was adapted from 'The Cell Cycle, principle of control' by Dr. D. O. Morgan [13].

The levels of cyclins periodically change between different phases of the cell cycle determine the assembly and activation of cyclin/CDK complexes, activating downstream biological processes via phosphorylation. In eukaryotic cells, four cyclins are heavily involved in the progression of the cell cycle, G_1-cyclin (Cyclin D), G_1/S-cyclin (Cyclin E), S-cyclin (Cyclin A), and M-cyclin (Cyclin B).

This regulatory network, composed of cyclin/CDK complexes, is therefore able to schedule the activities of cell cycle events. To control the activities of cyclin/CDKs complexes, CDK inhibitors (CKIs) negatively regulate the cell cycle by targeting CDKs. Further details of the biological events conducted by cyclins/CDKs will be introduced in following paragraphs.

G_0-phase

G_0-phase is a stage in the cell cycle where the cells are positioned in a quiescent state. The cells enter the G_0-phase in response to the shortage of nutrients or due to a lack of growth factor signalling. The quiescent stage is also defined as a reversible cell proliferation arrest where the cell cycle machinery is silent. Current studies indicate that a specific set of genes is up-regulated in a 'quiescent program', which ensure the cells remain in a non-dividing state but also contain the reversibility of the cell cycle arrest [22].

G_1-phase

G_1-phase is the first stage in the cell cycle. During this early stage in interphase, the cell increases its size and generates enough proteins, preparing for DNA replication and centrosome duplication for mitosis. During G_1-phase, cyclin D is synthesised, driving the transition between G_1- and S-phase. The cascade of cyclin D regulation is initiated by its binding to cyclin-dependent kinase 4 (CDK4) or cyclin-dependent kinase 6 (CDK6). The cyclin D/CDK4/6 complex is involved in the activation of transcriptional activator E2F, cyclin E, A, and cell division cycle 25A (CDC25A). The self-activation of E2F sets a positive-feedback loop, which enhances the activation of the cell cycle-regulated proteins mentioned above.

At the end of G_1-phase, CDC25A is involved in the activation of the cyclin E/cyclin-dependent kinase 2 (CDK2) complex, driving the cell cycle to pass through the restriction point (i.e. the first checkpoint in the cell cycle, see Figure 1.2), which results in an irreversible commitment to cell cycle entry. The level of cyclin E is reduced after the cell enters S-phase. Cyclin D also coordinates with cyclin E during the later stages of G_1-phase to regulate progression to S-phase.

Before entering the S-phase, DNA damage (if present) activates the CKIs and blocks the cell cycle at the G_1-phase checkpoint until the damage is repaired e.g. inhibitors of CDK4 (INK4) family members (i.e. p15, 16, 18, and 19) specifically target CDK4 and 6, blocking the cell cycle [23]. Cip/Kip family members (i.e. p16, p21, and p27) inhibit activities of all CDKs combined with cyclins [24].

The centrosome duplicates when a cell enters the cell cycle. The cyclin E/CDK2 complex not only triggers the transition between G_1- and S-phase but also initiates the centrosome duplication indirectly. In order to prepare for centrosome duplication, the polo-like kinase 2 (PLK2) catalyses the phosphorylation of protein nucleophosmin family member 1 (NPM1), shuttling it into the unduplicated centrioles, and licenses the development of cartwheel structure in G_1-phase [25].

S-phase

During S-phase, the cell cycle control system is heavily involved in chromosome [26,27], histone [28], and centrosome duplication [29,30], regulating by cyclin/CDK complexes. In eukaryotic cells, linear chromosomes are giant structures composed of DNA and proteins and their duplication is a complex biological event (Figure 1.4).

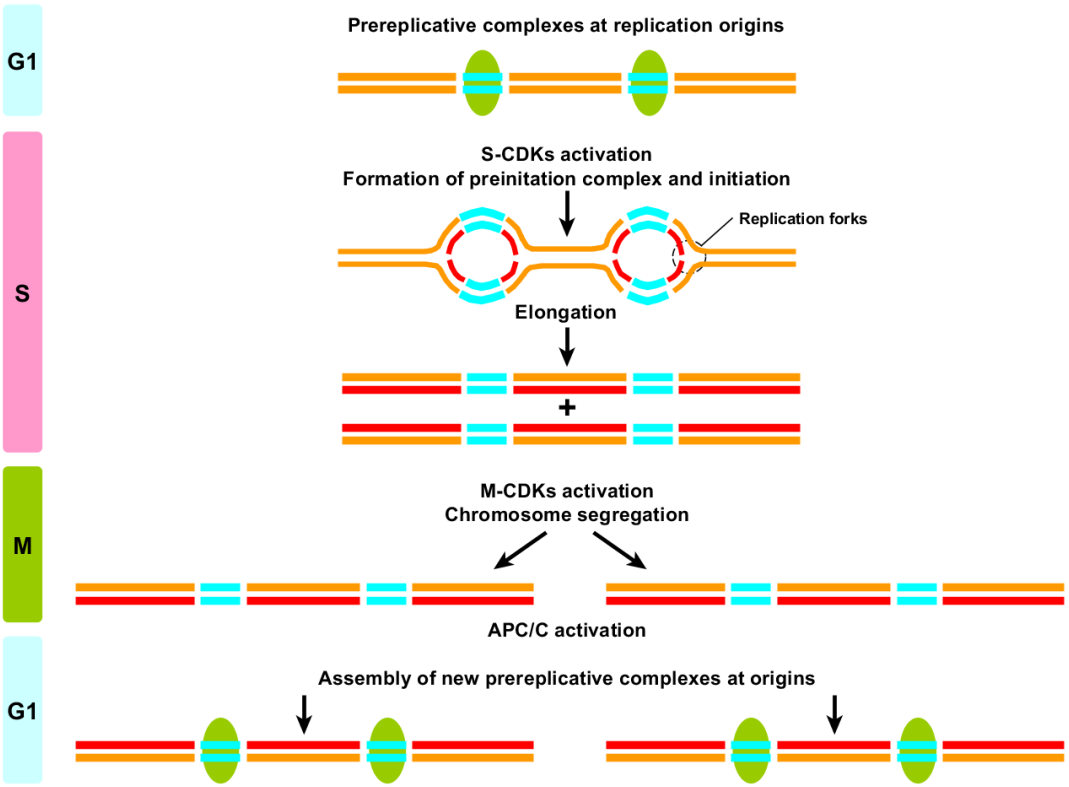


Figure 1.4: Cyclin conducted chromosome duplication in S-phase. Preparation for chromosome duplication is triggered in G_1-phase when pre-RCs (pre-replicative complexes; see green ellipses) are recruited to the origins of replication (see blue line). During S-phase, activity of S-CDKs mediates the development of preinitiation complexes by unwinding the dsDNA at origins. DNA duplication occurs at the moving replication forks. Duplicated chromosomes are segregated during M-phase. Blue lines represent the replication origins. Orange lines represent the original template strand. Red lines represent newly synthesized DNA. This figure was adapted from 'The Cell Cycle, principle of control' by Dr. D. O. Morgan [13].

The biological processes combined with the CDKs regulation are shown as a flowchart in Figure 1.4. In G_1-phase, the binding of pre-replicative complexes at multiple replication origins is preparation for the DNA duplication. During S-phase, activated S-phase CDKs conduct the formation of preinitiation complexes triggering the DNA duplication, following the replication forks. DNA duplication is completed by the end of S-phase as sister chromatids undergo condensation. In M-phase, activated M-phase CDKs conduct chromosome segregation until cell division is finished.

DNA replication starts at numerous 'origins of replication' on chromosomes, at specific DNA sequences with high AT content where replication is physically triggered [31]. To start DNA duplication, initiator proteins bind to the origin positions and helicase works to unwind the double helix of DNA. Duplication enzymes are then loaded onto the single stranded templates of DNA leading to the elongation phase (Figure 1.5) [32].

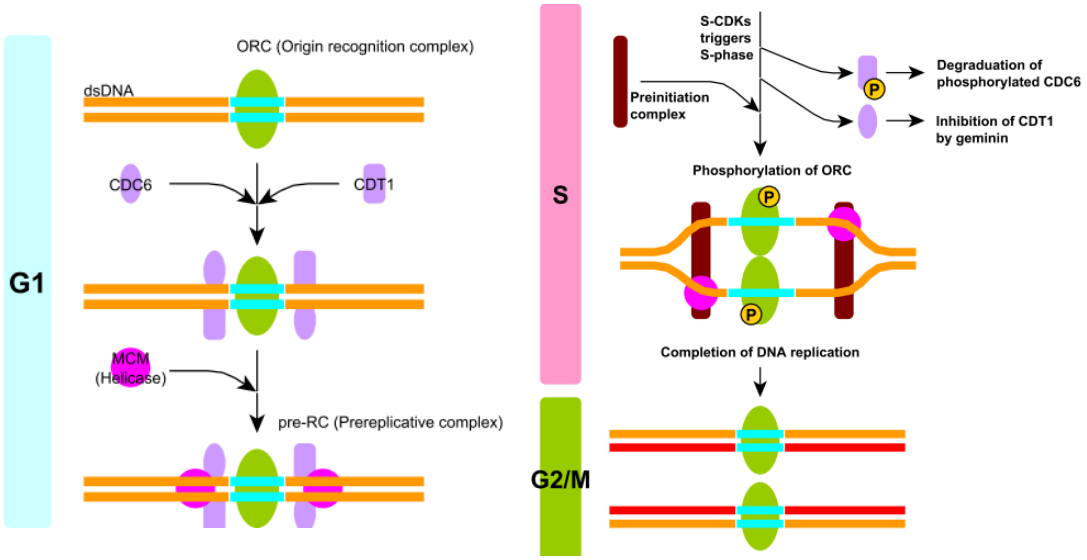


Figure 1.5: Initiation of DNA replication. At the early stage of the G_1-phase, CDC6 and CDT1 are both located near to the ORC and are then combined with the MCM resulting in the development of the pre-RC. During S-phase, the increase of S-CDK activity triggers the assembly of preinitiation complex binding at the origin of DNA replication. To avoid the reduplication of dsDNA, S-CDKs also targets and triggers the degradation of CDC6 by phosphorylation. In addition, CDT1 is inactivated by the protein geminin. This figure was adapted from 'The Cell Cycle, principle of control' by Dr. D. O. Morgan [13].

In order to avoid the re-duplication of chromosomes during the cell cycle, DNA duplication is divided into two major steps. Firstly, the proteins of pre-RC are recruited at origins of replication (Figure 1.5). The pre-RC recruitment not only licenses the initiation of DNA duplication but also guarantees the duplication only starts from the location, where the pre-RC is located. Secondly, the accumulation of pre-RC located on the DNA results in the formation of the pre-initiation complex (Figure 1.5). The DNA helix is then unwound by pre-initiation complex in preparation for the loading of DNA polymerases onto the single stranded DNA and to commence the DNA duplication. The pre-RC is then permanently dissociated from the binding sites on DNA without any reassembly (Figure 1.5) [32].

Figure 1.5 shows the regulatory network involved at the beginning of DNA duplication [33]. During G_1-phase, the ORC binds to the replication origins. CDC6 and CDT1 bind to a position nearby the ORC on DNA to form a large pre-RC complex and then recruits the minichromosome maintenance (MCM) complex. The binding of pre-RC licenses DNA replication to proceed. The MCM complex proteins are thought to unwind the DNA functioning as the major DNA helicase. This means that the aim of pre-RC is to recruit the helicase thereby playing a role in continuing the subsequent DNA replication process [34].

In S-phase, once the assembly of pre-RC is finished, DNA duplication is ready to start from the ORC's DNA. Activated S-CDKs that arise just after the G_1/S border recruiting a number of additional protein complexes to become a giant pre-initiation complex unwinding the DNA helix and starting DNA duplication. A high level of S-CDKs also induces the dissociation of pre-RC located at the origin.

Phosphorylation of the ORC and CDC6 by S-CDKs inhibit their function to initiate DNA duplication. In addition, anaphase progress complex (APC/C) is involved in the termination of pre-RC recruitment at the end of G_1-phase. Between the end of mitosis and early stage of G_1-phase, the protein geminin is destroyed by APC/C. The regulation between CDKs and duplication-regulated proteins (i.e. pre-RC) guarantee that DNA duplication can only be triggered once per cell cycle [32]. By the end of S-phase, duplicated DNA is undergoing the establishment of sister chromatid cohesion [35]. This is a process where the cohesin ring pairs sister chromatids together until metaphase, facilitating the attachment of mitotic spindles onto the chromosomes [35].

It is interesting that the centrosome and chromosome duplication share similar principles, in that they are semiconservative and controlled by S-phase cyclins/CDKs complexes [29,30]. In addition, the centrosome duplication is tightly controlled to allow only one duplication per cell cycle [36]. During S-phase, the activation of cyclin A and E in combination with CDK2 initiate centrosome duplication [29,37-41]. Duplication of the daughter centriole starts from the proximal end (cartwheel structure) to the distal end, which is surrounded by 9-fold symmetrical triplicate microtubules (Figures 1.6C and D). Matured daughter centrioles are surrounded by pericentriolar material (PCM). Owing to the complexity of the centrosome's structure and duplication, this is discussed in more detail in chapter 5.

G_2-phase

G_2-phase is the final stage of interphase and takes place before the cell enters mitosis. This ends at the early stage of prophase whilst the duplicated DNA is undergoing condensation, preparing the cells for sister chromatid segregation in M-phase. In preparation for chromosome segregation, a matured daughter centriole separates from the mother centriole, giving two centrosomes.

It is interesting that the transition between G_2-phase and mitosis is determined by the activity of cyclin B/CDK1 complex. Cyclin B expression increases at the end of S-phase just after

DNA duplication and accumulates in the cytoplasm throughout the G_2-phase whilst binding with CDK1. Chromosome duplication failure activates the p21 and blocks the cell cycle at the G_2/M checkpoint until the damaged DNA is repaired. The activated p21 targets the cyclin B/CDK1 complex in the nucleus [42].

M-phase

The cyclin B/CDK1 complex regulates the final stage of the cell cycle i.e. the mitosis. Firstly, the cyclin B/CDK1 complex plays a remarkable role to initiate the nuclear envelope break down (NEBD). Secondly, the cyclin B/CDK1 complex mediates the condensation of sister chromatids into compact, rod-like shapes. In addition, the cyclin B/CDK1 complex conducts the assembly of the mitotic spindle ensuring that the sister chromatids are equally segregated into two daughter cells.

In association with the cyclin B/CDK1 complex, the family of polo-like kinase (PLK) are involved in the regulation of the centrosome during M-phase. For example, CDK1 and PLK1 are both involved in the regulation of centrosomal protein 55kDa (CEP55) [43-45] and kinesin family member 11 (KIF11) [46]. CEP55 governs the physical division of cells during cytokinesis. KIF11 is a protein that moves along the microtubule (MT) filaments i.e. mitotic spindle, by using the energy from adenosine triphosphate (ATP). KIF11 regulates the extension of microtubules attaching to sister chromatids [46]. Both the cyclin B/CDK1 complex and centrosomes are important in determining a well-organized chromosome segregation during M-phase and cytokinesis. In M-phase, cyclin A is later degraded during metaphase of mitosis. Owing to the fact that the centrosome is heavily involved in the regulation of mitotic activities, the centrosome will be introduced first and then its role in mitosis.

The centrosome

The centrosome is known to be the centre of primary microtubule-organisation in animal cell and was first described at the end of 19^{th} century. The centrosome governs the motility, adhesion and polarity of cells during interphase. The centrosome also regulates the spindle assembly during mitosis. Failure in the organisation of the spindle pole results in many cancers owing to the genomic instability [47]. This is because extra or abnormal centrosomes can lead to unequal chromosome segregation [48]. Therefore, a tight control of centrosome life cycle determines a healthy cell proliferation.

The role and importance of the centrosome was discovered in the end of the 19th century by Boveri who sensed the connection between the number of centrosomes and cancer. Centrosome duplication during S-phase was then discovered in the 1980s [49]. To better understand the centrosome life cycle, centrosome development, duplication, and maturation is described here.

The major function of the centrosome is to conduct microtubule (MT) nucleation working as a microtubule organisation centre (MTOC). In connection with the centrosome by their negative end, MTs are able to provide the support for cytoskeleton influencing the cell polarity. In order to maintain the position of MT-connected organelles in the cells, as the supporter of MT, the centrosome is localised close to the nucleus.

The structure and localisation of the centrosomes has been studied with the help of electron microscopy (Figures 1.6A and B). A regular centrosome is composed of a mother and a daughter centriole surrounded by PCM, which is an electron-dense matrix. The mother centriole has distal appendages anchoring the centrosome at a fixed position on the cell membrane [50]. Subdistal appendages govern the nucleating of the microtubule network [51]. The matured centrosomes conduct chromosome segregation during M-phase [52].

Mitosis

Mitosis is composed of five stages according to the state of centrosome regulation including prophase, prometaphase, metaphase, anaphase, and telophase.

Prophase

Prophase is the stage where the sister chromatids are compacted, undergoing chromosome condensation. It is also the stage where the microtubules are assembled, conducted by the centrosome (Figure 1.7). In early prophase, the CDC25B and C are involved in the activation of cyclin B/CDK1 by removing the inhibitor WEE1 (i.e. G_2/M checkpoint kinase) and myelin transcription factor 1 (MYT1) kinase. Activated cyclin B/CDK1 is then involved in the initiation of NEBD and spindle assembly conducted by centrosomes.

During chromosome condensation, the condensins are activated by cyclin B/CDK1 complex to gradually compact a mass of partially catenated DNA into visible chromosomes [53,54]. Sister chromatids are attached to each other at the centromere, resulting in the X-shape (Figure 1.7A). Meanwhile, two centrosomes move to the opposite poles in the cytoplasm (Figure 1.7A). They initiate the microtubule assembly by recruiting the γ-tubulin, which becomes the mitotic spindle.

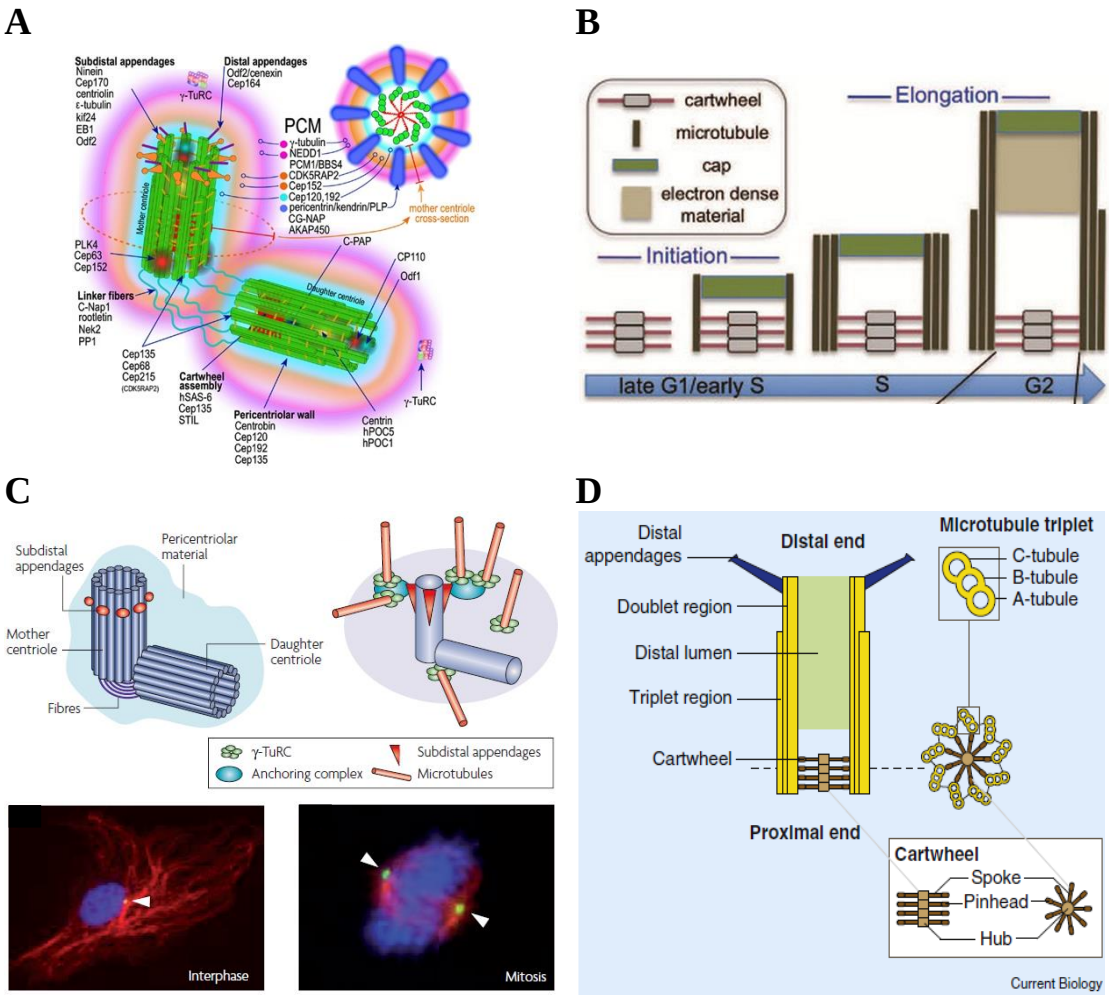


Figure 1.6: Structure of the centrosome and location in cells. A number of the best examples demonstrate the development of the centrosome and localisation in cells. **A)** The interphase centrosome including a mother and daughter centriole at longitudinal section. This figure is courtesy of Pihan *et al.*, 2013 [52]. **B)** The development of a mother centrosome is shown in a cartoon flow chart. This figure is courtesy of Kitagawa *et al.*, 2015 [55]. **C)** Cartoon graphic in 3D and images of protein localisation to demonstrate the structure of interphase and mitosis centrosomes. Chromosome is labelled with DAPI (blue). F-actin is stained by phalloidin (red) colour. GFP signal is the centrosome where the arrow pointed. This figure is courtesy of Fukasawa *et al.*, 2007 [56]. **D)** Longitudinal and cross section of the mother centrosome structure. This figure is courtesy of Azimzadeh *et al.*, 2010 [57].

Prometaphase

Prometaphase is the second phase of mitosis where the nuclear envelop dissociates into numerous ‘membrane vesicles’. It is also the stage where the mitotic spindles attach to the outer kinetochores on the sister chromatids (Figure 1.7B). The outer kinetochore is the area where the mitotic spindle connected with. The inner kinetochore is strongly connected with the centromere DNA, which is a special form of compacted chromatin [58].

Metaphase

Metaphase is the third stage in mitosis, where chromosomes are at the most condensed stage. During this stage, the mitotic spindle emerges from two centrosomes and attaches the kinetochore at opposite sites. Chromosomes are aligned at the equator of the mitotic spindles, which are equally positioned between the two centrosomes (Figure 1.7C). This is due to the equal pulling force from opposite mitotic spindles.

The metaphase to anaphase checkpoint senses whether the mitotic spindles have properly attached to chromosomes (Figure 1.7C). Only when all chromosomes are properly attached to mitotic spindles and are aligned in the middle between the two centrosomes, will the checkpoint allow the transition between metaphase and anaphase.

The cyclin B/CDK1 complex activates the securin, which inhibits the separase activity by binding with it. Without the targeting from separase, the cohesin ring is able to pair sister chromatids together until the end of metaphase. The increase in Ca^{2+} activates the Ca^{2+}/ calmodulin-dependent protein kinase II (CaMKII), mediating the activation of APC/C. Activated APC/C is then involved in the deactivation of cyclin B/CDK1 complex by degrading cyclin B, leading to the breakdown of the cohesin ring, preparing for chromosome segregation in anaphase [59]. The cyclin B/CDK1 complex is also involved in the metaphase arrest, ensuring the sister chromatids are properly attached to the mitotic spindle.

Anaphase

Anaphase is the stage where sister chromatids are segregated to the opposite poles of a cell. In addition to the regulation of cyclin B, the APC/C is also involved in the disassociation of sister chromatid cohesion by destroying the securin [60]. Paired sister chromatids, attached with the mitotic spindles, are pulled to opposite poles after the cohesin ring is dissociated abruptly (Figure 1.7D).

Telophase

Telophase is the final stage in mitosis where the interphase nuclear envelope is reformed and a mitotic cell is physically divided into two daughter cells. Materials for reshaping of the nuclear envelope are provided by parts of the endoplasmic reticulum [61]. The nucleoli also reappear following the reforming of the interphase nuclear envelope (Figure 1.7E). Once the reforming of the nuclear envelope is completed, the mitotic spindles start to be disorganized (Figure 1.7E) [61].

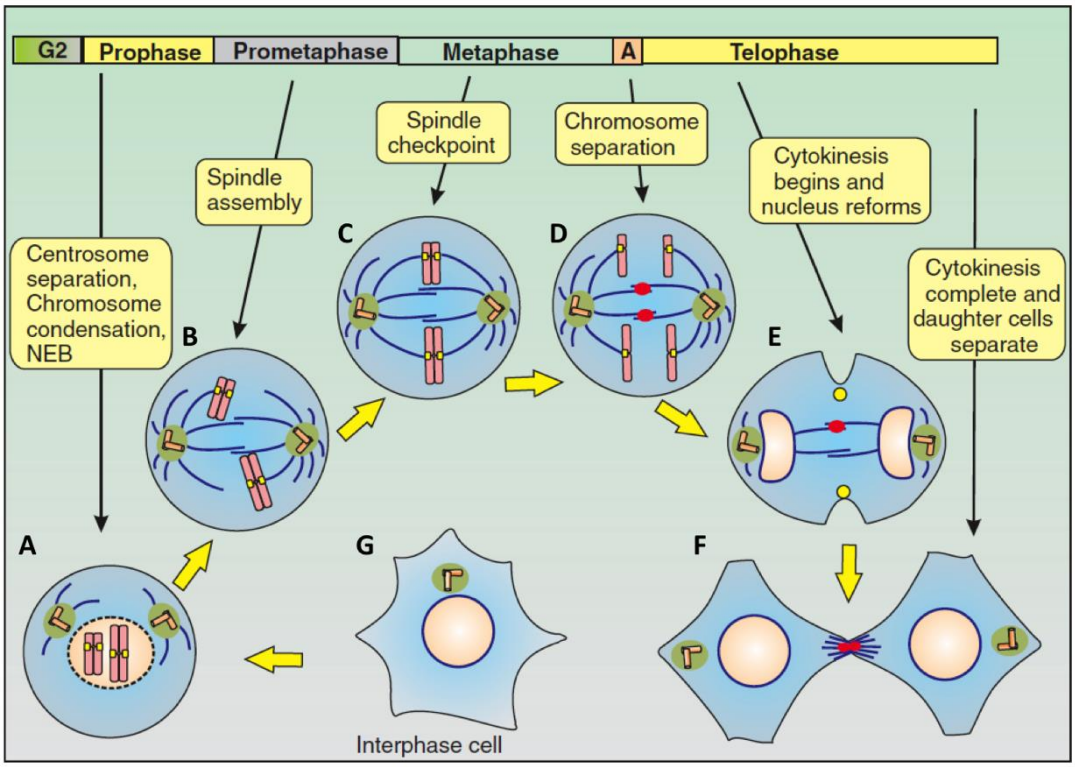


Figure 1.7: Summary of the morphological changes of centrosomes in mitosis. The centrosome is heavily involved in the M-phase regulation. **A)** The mother and daughter centrosomes are separated moving to opposite poles in early prophase. **B)** In prometaphase, the centrosomes initiate the spindle assembly and extension. **C)** In metaphase, the spindle checkpoint sense the proper attachment between the aligned sister chromatids and mitotic spindle, determining the transition between metaphase and anaphase. **D)** Anaphase chromosomes are pulled to opposite poles, whilst the mitotic spindles are contracting. **E)** The contracting of contractile ring physically **F)** divides the cell into two daughter cells, whilst the nucleus reform starts. **G)** The cytokinesis is completed, resulting in interphase cells. This figure is courtesy of Berridge *et al.*, 2014 [62].

Cytokinesis

Once mitosis is finished, cytokinesis ensures the equal separation of cells. Cytokinesis initiates with the development of the contractile ring, which leads to the accumulation of cleavage furrow. The cleavage furrow is the final connection of the cell membrane between two daughter cells undergoing cell division (Figure 1.7F).

The contractile ring is attached to the cell membrane. Before two daughter cells are completely separated, a structure called the midbody maintains the final association between the two daughter cells (Figure 1.7F). In the last stage of cytokinesis is abscission disassociates the cell membrane connected by midbody. CEP55 is activated by CDK1 to conduct the abscission and separate the cell into two daughter cells [63,64].

In summary, during cell proliferation and division, a large number of genes are periodically expressed to regulate this complicated regulatory network. By the end of the 1990s, array technologies were developed to monitor cell cycle-associated genes in parallel. Hence microarrays have played an important role in the study of the cell cycle at the transcriptional level. The history of array technology application will be discussed in following paragraphs including approaches used to identify the cell cycle-associated genes.

Transcriptomic analyses of the cell cycle

The availability of large-scale genetic information from whole genome sequencing technologies revolutionised the study of gene expression from the single gene to whole transcriptome level. This information was used to develop microarray technology. Microarrays evolved from a technique called Southern blotting and the first microarray was made by attaching an array of microscopic DNA spots to a solid face [65]. Microarrays can be used to monitor the expression level of huge number of genes at once. The probe can include a short sequence of a gene, which hybridises with a target (cDNA) [65]. The cDNA is reverse transcribed from isolated mRNA from cell or tissue samples. The hybridisation of the probe and target generates a signal, which is detected by a digital optical sensor [65]. The strength of the signal is used to determine the quantity of cDNA sequences in the target and consequently the level of expression of the relevant gene. For further detail of the history of the array technology development will be shown in the following paragraphs.

A brief history and principles of microarrays

Microarray technology

In 1975, the first array was created by plating *E. coli* containing different clones in an array format on agarose plates before transferring to a nitrocellulose filter; transferred clones were lysed releasing denatured DNA which was used to be hybridised with radioactively labelled RNA [66] (Figure 1.8A).

In 1979, the array production method was improved using robotic arms to spot a total number of 1,728 *E. coli* on to a 26 x 38 cm microtiter plate to increase the density of spots (Figure 1.8B). The concept of spotting techniques on filter papers was developed further in following years. For example, in 1980, the filter based array was used to identify the transcriptional activity of genes cloned in plasmids [67]. In 1986, the technique was used to study of single nucleotide polymorphisms (SNPs) by hybridising with chemically synthesised oligonucleotides, designed to perfectly match specific target sequences [68].

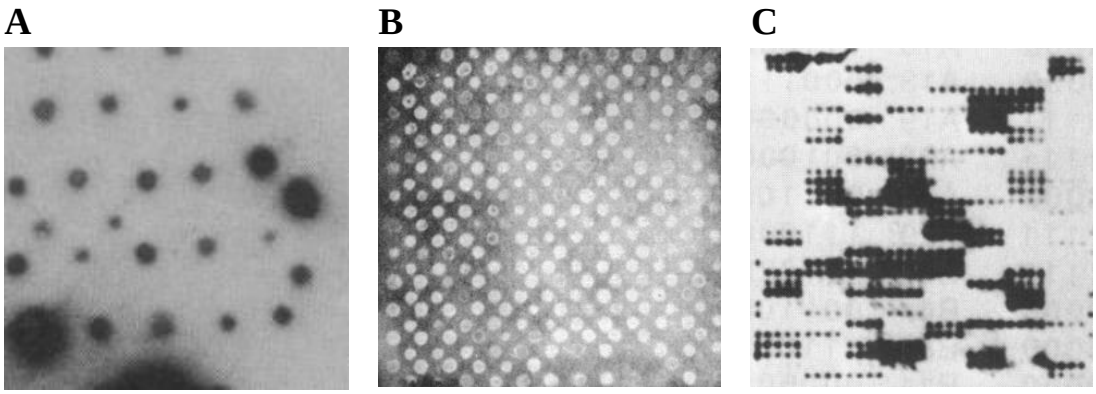


Figure 1.8: Radioactive monitoring of the RNA hybridisation with DNA probe on filters. A) A handmade array by Bumgarner *et al.*, 2013 [66]. **B)** A machine printed array by Gergen *et al.*, 1979 [69]. **C)** Assembly line manufactured array by Craig *et al.*, 1990 [70].

In 1990, an assembly line was used to generate filter based arrays in large quantity, commercialised robot arms were used to spot *E. coli* from microtiter plates to filters (Figure 1.8C) [70]. In 1992, oligonucleotide DNA probes were directly attached a glass support and facilitated more sensitive detection [65].

Since the first array was created in the 1970s, the development of DNA array technologies has evolved rapidly. For example, the oligonucleotide sequence can be designed to target specific regions of the genes of interest, enhancing the accuracy of probe hybridisation [66]. Array technologies can be summarised into three basic categories from the 1990s to the early 2000s and are discussed below.

Spotted arrays

The spotted array consists of synthesised oligonucleotides or PCR products used to target a set of genes of interest spotted on a chip at high density (Figure 1.9A). The spotted array allows flexibility and the array could be customised according to the experimental design. This technology allowed a very high density of probes to be spotted on poly-lysine coated glass supports by using slotted pins, these are similar to the structure of fountain pens (Figure 1.9B) [71]. Just by a single dip of a pin into a given DNA solution, this system can equally spot it on multiple arrays, allowing samples to be fluorescently labelled (Figure 1.9A).

In comparison with the radioactive labelling system, fluorescent labelling has a number of advantages. First, the fluorescent signal released is highly sensitive as the signal has a fairly large dynamic spectrum from perfect match to mismatch in hybridisation. Second, the cost

and complexity of the fluorescent labelling system is less expensive than the radioactive labelling system.

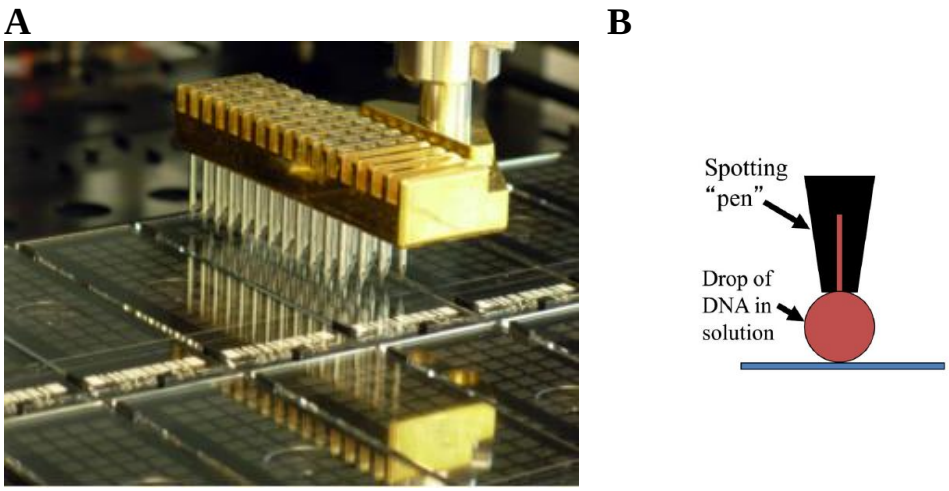


Figure 1.9: Fabrication of the spotted array. A) Chemically synthesised probes were spotted on the solid surface of an array prior to the hybridisation of samples. This figure is courtesy of The Scripps Research institute **[72]**. **B)** Probes are spotted on the solid face of an array. This figure is courtesy of Bumgarner *et al.*, 2013 [66].

In-situ synthesised arrays

The in-situ synthesized array is produced using modern photolithographic technologies where the probes are generated in parallel using a number of approaches, these include solid-phase synthesis [73], photolabile protecting groups [74], and photolithography [75] to conduct the chemical synthesis [74]. By combining these 3 technologies, light directed chemical synthesis probes can be directly generated on the array (Figure 1.10) [74].

Encouraged by the successful approach in 1991, the company called Affymetrix was founded by Dr. Stephen Fodor. In 1994, the in-situ synthesized array was commercialised by Affymetrix, demonstrating that a total number of 256 unique octa-nucleotides can be fabricated on a 1.28 cm^2 array [76].

The major advantage of modern photolithographic techniques is that the DNA probes are directly synthesised on the solid face of an array, allowing the fabrication of a complex array with a high density of probes (Figures 1.10A and B). In contrast, the disadvantage of spotted arrays is that a large number of cDNAs or PCR products need to be generated, purified, quantitated, and catalogued prior to probe spotting [77]. In addition, if some of the probes are generated from a cDNA library at low quantity, additional normalisation may be required to avoid the redundant spotting of cDNAs.

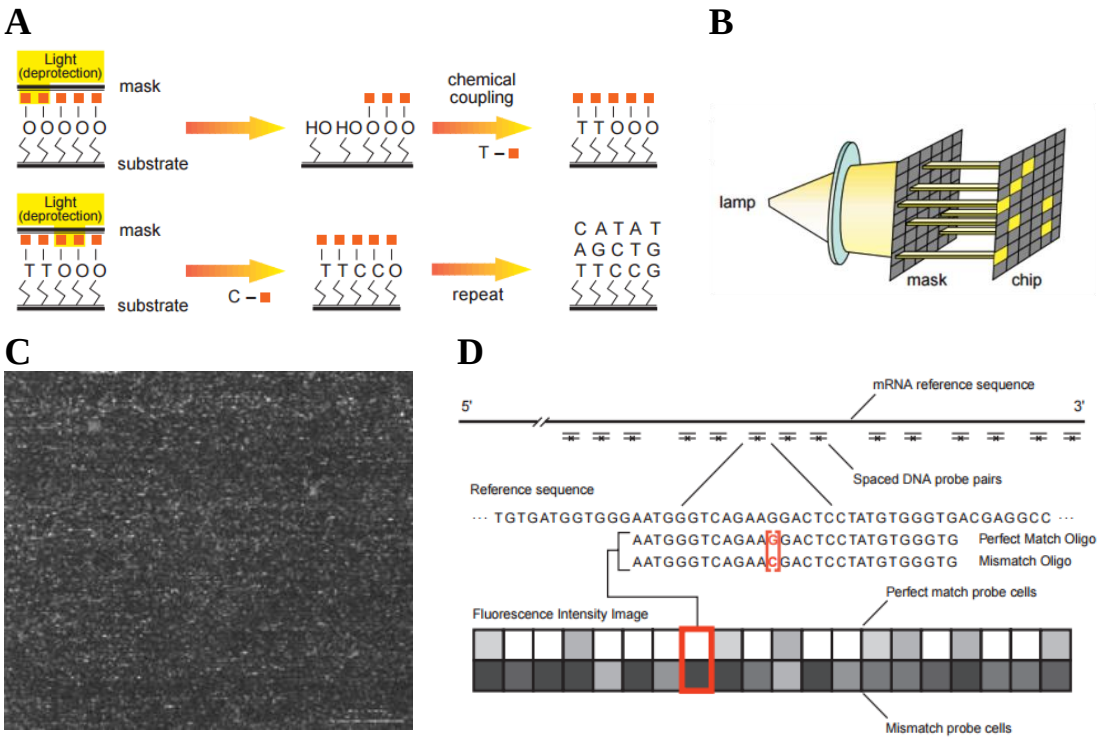


Figure 1.10: Modern photolithographic techniques. A) The processes of light directed oligonucleotide synthesis. The photolabile protection provided by the mask determines the sequence of probe synthesis. **B)** Schematic representation of the array-making device, including the lamp, mask, and array (chip). **C)** Gene expression level monitored by an in-situ synthesised array. **D)** The design of array and probes. The probes are chosen from the 3' end of messenger RNA (mRNA) (i.e. near the poly A tail), which avoids the selection of partially degraded mRNA. Signals from perfect matches (PM) and mismatches (MM) are used to increase the quantitative accuracy. These figures are courtesy of Lipshutz *et al.*, 1999 [75].

However, there are disadvantages when using spotted arrays, the probes are attached onto the charged lysine side chain of the spotted array by an interaction with the negatively charged phosphate back-bone of cDNA, and this interaction is thought to minimise the conformational freedom of attached cDNA probes, potentially reducing their affinity for complementary targets [77].

In the family of in-situ synthesized arrays, each array product requires a unique set of photolithographic masks to fabricate the DNA probes by directing the light to the array during different steps. As some of the probes are redesigned, the entire set of photolithographic masks must be reproduced, which increases the cost of production and limits the flexibility of each model of array.

In 2002, the micro-mirrors technology was developed by Nimblegen Systems Inc., which directs the light at pixels to conduct mask-less fabrication of probes on the array, using

digital light processor (DLP) to project a pattern of UV light (Figure 1.11) [78]. This digital mask provides another solution to generate the array without expensive and time-consuming photolithographic masks [78].

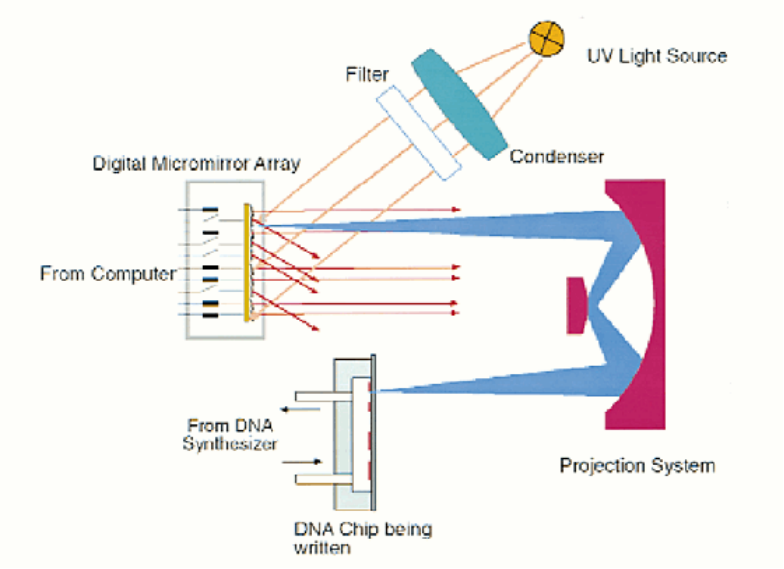


Figure 1.11: Schematic of the digital light processor (DLP). This is an illumination system with a UV light source. Filtered UV light is projected to the digital micro-mirror array, where the computer determines the sequence for oligonucleotide synthesis. The projection system accepts the UV light from the digital micro-mirror array and projects to the solid support of the DNA chip (microarray) and conducts the probe fabrication on the array. This figure is courtesy of Singh-Gasson *et al.*, 1999 [79].

The GeneChip developed by Affymetrix is the best example of signal channel arrays. The GeneChip is used to detect gene expression level by using a single labelling scheme for cDNA samples. For each probe set, there is a collection of probe pairs targeting a gene of interest. In general, a probe set includes 11 to 20 probe pairs, and each probe is a 25-mer oligonucleotide. Each pair of probes contains a complementary sequence to match the gene of interest perfectly, which is also called PM (Figure 1.10D). In contrast, another pair of the probes are designed as a control, which is used to discriminate non-specific match i.e. MM (Figure 1.10D).

Inkjet printing arrays

In 1996, the innovating inkjet printing technology allowed direct conduction of the synthesis chemistry on the solid face of arrays [80]. The inkjet printer is used to place nucleoside phosphoramidite (i.e. nucleotides used to synthesise probes) on the array, where the solid face of the array is pre-patterned with hydrophilic and hydrophobic regions (Figure 1.12).

The hydrophilic region allows the coupling of phosphoramidites dropped by inkjet printers to synthesize oligonucleotide probes (Figure 1.12).

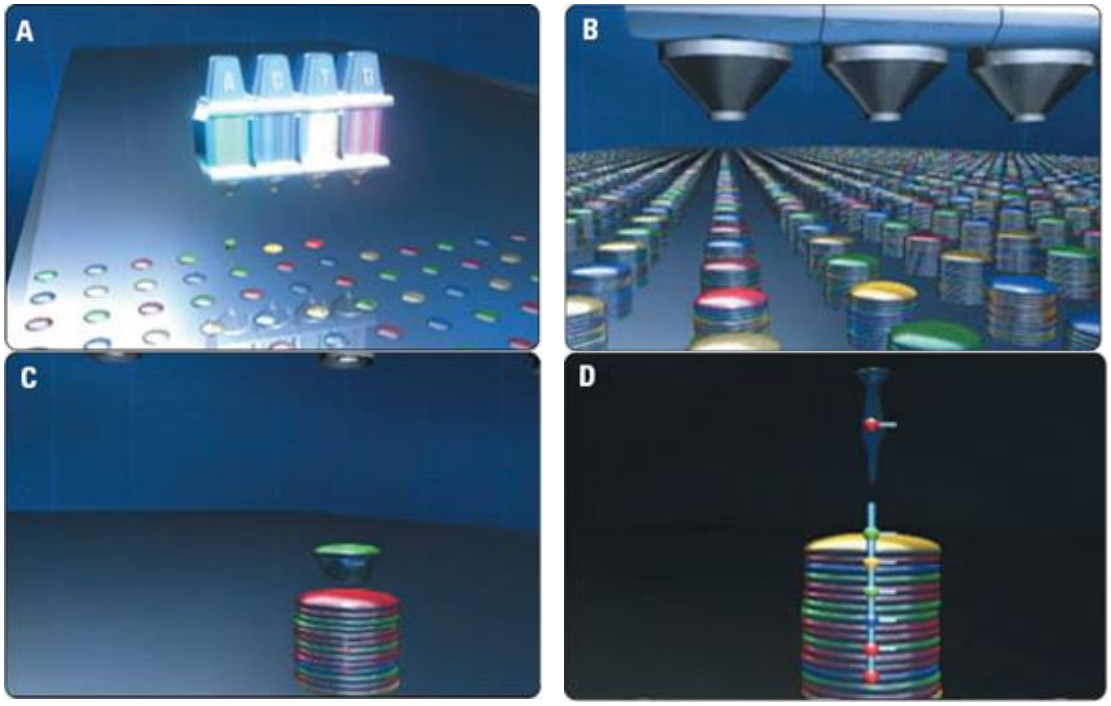


Figure 1.12: In-situ synthesis of the printing process of inkjet arrays. A) First layer of nucleotides are printed on the hydrophilic section on the array solid surface. **B)** Multiple layers of nucleotides are printed to fabricate the DNA probes. **C)** and **D)** show the transition between one nucleotide being added to the chain to induce chemical synthesis. These figures are courtesy of Agilent microarray technology [81].

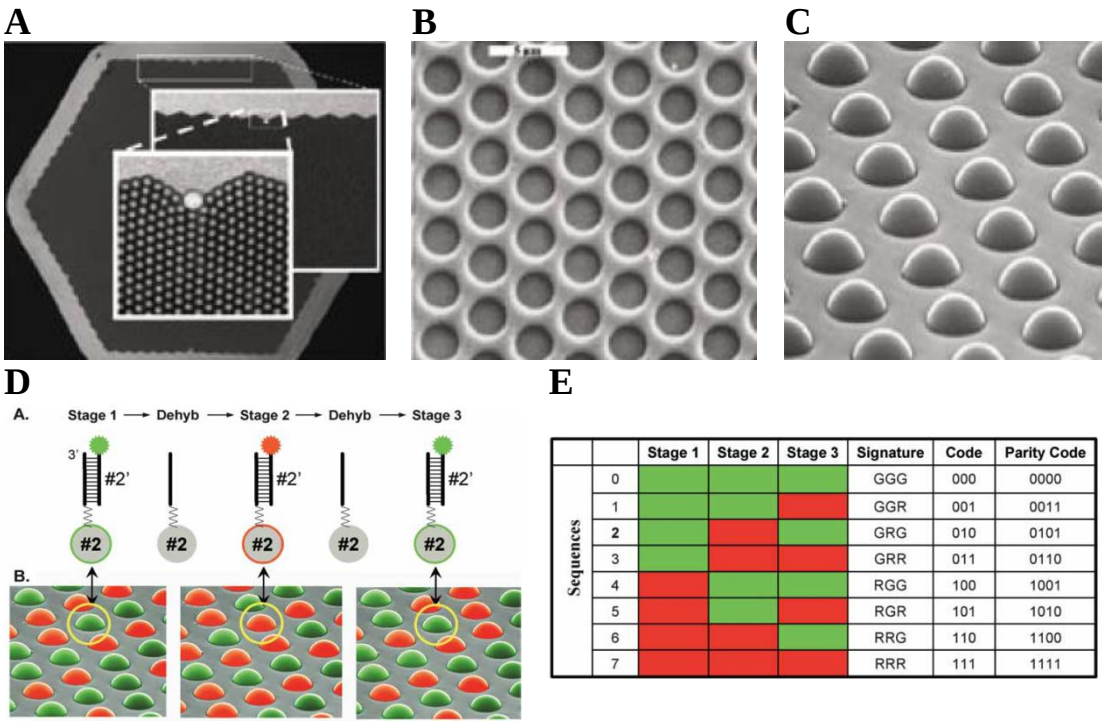
This technology was eventually improved, commercialised, and licensed to Agilent Technologies in 2001 [82]. The inkjet array shares the advantages of the Affymetrix arrays where they both directly synthesis the oligo nucleotides on the solid face directly. This technology was used to fabricate the probes for their array products until present.

Self-assembled arrays

In comparison with the array technologies discussed above, an alternative design of the array construction (i.e. the self-assemble array) was created by a group based in Tufts University lead by David Walt [83-86]. This technology was later licensed to Illumina.

The self-assembled array is a technique that attaches the DNA probes onto small polystryrene beads. These beads are deposited on wells, created by etching the end of optic fibres and assembled together as a fibre optic array (Figures 1.13A, B, and C) [87]. The fluorescent signal of hybridisation can be conducted by optic fibre from the array to the

distal end where the fluorescence microscope magnifies the signal and projects onto a charged-couple device (CCS) camera [87].



		Stage 1	Stage 2	Stage 3	Signature	Code	Parity Code
Sequences	0				GGG	000	0000
	1				GGR	001	0011
	2				GRG	010	0101
	3				GRR	011	0110
	4				RGG	100	1001
	5				RGR	101	1010
	6				RRG	110	1100
	7				RRR	111	1111

Figure 1.13: Structure of typical self-assembly arrays, labelling of beads in early and later version arrays. A) The structure of assembled optic fibres, containing 50,000 to 60,000 fibres. **B)** Etched end of an array of optical fibres. **C)** Beads loaded into wells. **D)** Localisation of randomly loaded beads by a sequential hybridisation process is demonstrated by taking a single bead as an example. The fluorophore signal is switched to different colours at different stages, resulting in **E)** a combination of binary digit code (i.e. 010101). Each unique combination of binary digit code represents a specific probe, providing the information for the target DNA samples. These figures are courtesy of Walt D.R. *et al.*, 2006 and Gunderson *et al.*, 2004 [87,88].

In early versions of this technology, the position of each probe was identified by labelling with different fluorophore combinations (i.e. fluorescent chemical compounds that can release signal by light excitation) [83-86]. However, this labelling system limited the maximal number of unique beads that can be localised. Therefore, the later version of the self-assembled arrays localise each unique bead by decoding the signal flash of fluorescently labelled oligonucleotide tags, which are sequentially released in a series of steps with different combinations (Figures 1.13D and E) [88].

Testing cell cycle synchronisation before studying transcriptional regulation in the cell cycle by microarray technologies

The transcriptional and post-transcriptional regulation of genes is important in the regulation of the cell cycle. Using synchronisation methods that collect cells at the same phase of the

cell cycle, we can investigate particular biological events and their consequences to cell proliferation. This means that to identify the genes that are up-regulated in the cell cycle, cells must be synchronised before their periodical expression levels can be monitored by microarray [89]. Many studies have investigated cell cycle regulation in this way including in budding yeast [90,91], fission yeast [92], Gram-negative bacteria [93], HeLa cells [94], mouse embryonic fibroblasts [95], and human fibroblasts [22,96-98].

Array analyses can be used to identify cell cycle-associated genes in human cells

By investigating transcriptional control we can determine how cell cycle-associated proteins are regulated. The first microarray analysis used to identify human cell cycle-associated genes was performed in 1999 (Table 1.1) [96]. This pioneering work utilised the serum refeeding technique, validating the methodology and efficiency of this technique for investigating transcriptional regulation of the cell cycle.

Cluster analysis of human fibroblast cells suggested that 517 of 8,613 human genes represented on the array were up-regulated in response to the refeeding of serum [96]. 224 of the 517 genes were previously uncharacterised genes suggesting they may have a putative function in response to serum refeeding (Table 1.1). Of the remaining genes with known functions, 25 cell cycle-regulated genes were identified.

In addition, genes encoding G_0-CDKs that inhibit cell proliferation, were shown to be up-regulated in samples of cells harvested at a quiescent stage [22]. These findings validated that serum deprivation can block cells at G_0-phase. After serum refeeding, up-regulation of cyclin and CDK genes encoding protein from the S- and M-phase indicated that serum deprived fibroblasts respond as expect to serum refeeding.

The success of the first microarray analysis in identifying characterised and uncharacterised genes with a role in human cell proliferation gave rise to numerous subsequent studies (Table 1.1). Cho RJ, *et al.*, 2001 used double thymidine-blocks to halt human fibroblast at S-phase temporarily, before releasing them to generate a large population of synchronised cells [97]. They also used enrichment analysis to classify the biological function of cell-cycle associated genes. Of 6,800 genes monitored by microarray, 731 genes were clustered with cell cycle-regulated genes according to their expression pattern. For 344 of 731 genes, little functional information is available (Table 1.1) [97].

In another study from 2002, HeLa cells were used to identify genes that are expressed during the cell cycle (Table 1.1) [94]. In this experiment, HeLa cells were synchronised by three

different techniques; 1) double-thymidine block, 2) thymidine-nocodazole block, and 3) mitotic shake-off. The thymidine-nocodazole and mitotic shake-off arrested the cells at M-phase while double thymidine arrested the cells at G_1/S border. 874 cell cycle-associated genes were identified. 194 of identified genes were previously uncharacterised (Table 1.1). Their results also indicated that the genes involved in DNA duplication and sister chromatids segregation were highly up regulated in tumours [94].

While some groups worked on the identification of genes up-regulated between the different phases of the cell cycle, other research groups focused on the regulation of genes that are up-regulated during the quiescent stage e.g. Coller, *et al.*, 2006 (Table 1.1) [22]. They investigated the expression patterns of genes that are up-regulated at G_0-phase. Three different techniques were used to arrest the fibroblasts at a quiescent stage; 1) mitogen deprivation (removing of platelet-derived growth factor in cell culture arrested cells at a quiescent stage), 2) high cell density (contact inhibition), and 3) disruption of cell adhesion (culture plates coated with agarose to inhibit cell adherence). Interestingly they found that 95% of DNA synthesis was decreased during the quiescent stage. They found that quiescence was not a prolonged pause in G_1-phase. Up- and down-regulation of genes from array analysis suggested that quiescence is an activated stage in the cell cycle. In comparison with the control, they found a small set of genes that were up regulated across all three conditions that arrested the cells at a quiescent stage [22].

To identify any overlap in the list of cell cycle-associated genes in fibroblasts synchronised either by serum deprivation or by thymidine-block, an array experiment was performed by Bar-Joseph, *et al.*, 2008 (Table 1.1) [98]. In order to categorise the core set of cell cycle-associated genes, they also compared their list of cell cycle-associated genes with the genes identified in previous research using HeLa cells in 2002 [94]. They identified 480 genes that were periodically expressed after arrested fibroblasts re-entered the cell cycle. Of these genes identified in fibroblasts, 362 genes were found both up-regulated in HeLa and fibroblast cells.

However, the array that Bar-Joseph, *et al.*, 2008 used only covered two thirds of genes (i.e. Affymetrix® Human Gene U133A 2.0, see Table 1.1) [98]. Therefore, a complete human gene dataset and the latest array technology can provide us more information of cell cycle-associated gene expression, which was utilised in this thesis (i.e. Affymetrix® Human Gene ST 1.1 Arrays can monitor 21,014 gene expression) [22,96-98].

Year	Microarray platform	Well-characterised human genes	Clustered Genes	Channel	Cell type	Synchronisation method	Block phase	Ref.
1999	Affymetrix (Customised spotted array)	8,613	224 of 517 clustered genes were unknown	Dual	Human fibroblast	Serum deprivation	G_0	[96,99]
2001	Affymetrix (1st generation oligo array)	6,800	344 of 731 clustered genes were unknown	Single	Human fibroblast	Thymidine	S	[77,97]
2002	Customised spotted array (Stanford Microarray Facility)	16,332	194 of 874 clustered genes were unknown	Dual	HeLa	Thymidine	S	[94]
						Thymidine + Nocodazole	M	
						Mitotic shake off	M	
2006	Affymetrix (HuGENE-FL)	5,600	128 genes regulated by specific arrest signals	Single	Human fibroblast	Mitogen withdrawal	G_0	[22]
						Contact inhibition		
						Loss of adhesion		
2008	Affymetrix (U133A 2.0)	14,500	362 genes both up-regulated in fibroblast & HeLa	Single	Human fibroblast	Serum deprivation	G_0	[98]
						Serum deprivation + Thymidine block	S	

Table 1.1: Summary of the array experiments used to study the human cell cycle-associated transcriptional network.

Microarray and bioinformatics

In 1986, the human genome project was launched to fully sequence the human genome [100,101]. In 2003, the completion of the human genome project profoundly benefited the development of microarray. Scientists could now turn their attention to study the regulation of biological events at the transcriptional level e.g. the cell cycle. The application of these array approaches was one of the earliest examples where biologists entered into the big data era e.g. the gene expression activities in the human cell cycle [22,96-98,102,103].

Many statistical approaches were developed to minimise the diversity of raw array datasets as the hybridisation between probes and samples can be considered as numerous interactions at the molecular level. The aim of a typical array data mining analysis is to identify genes that are statistically significant i.e. up/down-regulated at a fixed time point. Further approaches can be used to cluster sets of genes according to their expression levels across a number of time points. For each particular cluster, clustered genes with previously uncharacterised functions can be investigated according to additional analyses e.g. loss of function analysis (i.e. gene silencing) during cell division, subcellular protein localisation (i.e. predicts a putative role in the biological events based on the location in the cells). Before carrying out any of these analyses, the question of normalisation and data transformation of the array dataset has to be addressed. Various statistical techniques will be discussed in the following paragraphs.

Equalise the diversity of array datasets

Before we start to identify genes of interest, a number of questions need to be asked: 1) are probe intensities across multiple arrays comparable? 2) if not, does the experimental variability affect the distribution pattern? Statistical tools can be used to equalise the variability within array data or across an array dataset i.e. normalisation.

The aim of normalisation is to remove the impact of non-biological effects in an array dataset, including experimental artefacts and random noise that are due to the technical and systematic variation [104]. Owing to the different array structure and labelling system, the normalisation for single channel array (e.g. oligo arrays from Affymetrix GeneChip) is different from dual channel array (e.g. spotted arrays).

For single channel arrays, an array is only hybridised with one sample. In contrast, the same gene from different samples are hybridised with the same spotted array. Therefore, the aim of normalisation for single channel array was developed for between-array normalisation i.e.

robust multi-array average (RMA) normalisation [105]. The spotted or cDNA array associated normalisations are designed for within and also cross array normalisation.

Using the Affymetrix GeneChip as an example for RMA normalisation, there are two steps in single channel array analysis, i.e. background correction and normalisation of array data. RMA uses PM probes to conduct the background correction. A second step, RMA conducts the quantile normalisation at probe level throughout an array dataset, the mean quantile is calculated to replace the raw value in the original data [106]. Normalised array data are ready for data mining.

Data mining

Array datasets are usually very large as the number of variables influences the analytical precision. Therefore it is very useful to narrow down the number of genes of interest as they are distinguished to the controls, which is based on the fold change of expression level (e.g. two-fold). There are several commonly used statistical methods used to filter the genes e.g. t-test [107].

For the mining of gene expression data, cluster analysis appears to be one of the most popular methods, which can be used to find a set of genes, which are co-regulated and functionally related [108,109]. As an unsupervised approach, the clustering can divide entire data into groups where clusters of genes or samples have similar expression profiles e.g. S-phase genes that are up-regulated at a specific time point in the cell cycle are clustered with genes whose roles are previously uncharacterised.

There are two widely used clustering methods i.e. hierarchical and k-means. Hierarchical clustering is an unsupervised approach that places genes together where clustered genes share similar expression profile [110]. The relationship between each sample in data can be estimated quickly and shown using a large-scale tree diagram. In contrast, K-means analysis is a non-hierarchical clustering method. K-means clustering undergoes an iterative process i.e. the re-evaluation is repeated to define the current centre in each existing cluster until the desired number of clusters is reached [111]. During the process of re-evaluation, a sample could be clustered in different clusters at different times [111]. Clusters are defined using Euclidean distances to minimise the difference between clustered samples but increase the difference between clusters [111].

For the last few decades, traditional statistical analyses are normally used to determine the effects of one or two conditions on a given set of samples and require multiple replicates of

the experiment to be done. However, for expensive analysis, such as array experiments, multiple repeats may become inherently expensive and traditional statistical analyses may be impractical for large-scale screens or analysis with various conditions.

Gene co-expression network analysis

In large-scale data mining, traditional analysis can only determine the level of an effect according to conditions in that experiment but are unable to explore the relationship between different samples i.e. the p-value of a t-test cannot be used to group genes, which have similar expression profiles. However, correlation based techniques can classify samples into multiple clusters e.g. the similarity between samples in hierarchal clustering can be determined by the Pearson correlation coefficient.

Network analysis has been proved to be capable of analysing biological data [112,113]. However, it has been rarely used over the past few decades owing to a lack of supporting software and powerful computers. Only recently has software with a friendly interface been developed to allow biologists to analyse array datasets, e.g. BioLayout *Express*^{3D} [113].

An array data set is a collection of transcriptional measurements from thousands of genes [112]. To demonstrate the level of similarities of a particular gene's expression profile between other genes, the network analysis uses edges to connect a node (i.e. each node represents a gene of interest) to others and creates a network. The concept of building a gene-expression network is described as follows.

To investigate a gene's expression profile within an array dataset in this thesis, I have performed network structure analyses. The networks are composed of nodes and edges. The strength of an edge represents the similarity between gene expression as represented as nodes and was determined by calculating pair-wise Pearson correlation matrix between each transcript on an array (Figure 1.14). The coefficient value is between +1 and -1 indicating that they are either positive or negative correlation, where 1 is total positive correlation, 0 means no correlation, and -1 equals total negative correlation. Figure 1.14 demonstrates a simple network structure built based on the correlation coefficient that defines the similarities between different transcripts (using edges weight colours to demonstrate the similarity).

Owing to the need for large-scale data mining at a systems level, BioLayout *Express*^{3D} was developed to visualize data obtained through network structure based analysis, by clustering the samples from a given biological data set [112]. Generally, the maximal capacity for data

visualization of a network is approximately 45,000 nodes, which are connected by more than $5x10^6$ edges. MCL (Markov CLustering) algorithm is used for network clustering analysis.

The Markov Cluster Algorithm (MCL) is an unsupervised cluster algorithm simulating the stochastic flow in a network structure i.e. a network composed of nodes and weighted/unweighted edges (Figure 1.14) [114]. This simulation is based on two operations (i.e. expansion and inflation) to identify clusters in which samples with similar features are placed together [114]. Expansion involves modelling the flow spreading out through the entire network. Inflation determines the contraction of flow in the network, making high/low current of flow in different network domains [114]. Taken together, undergoing the MCL simulation, the current of flow is high in a number of specific network domain but low between each other, determining the number of clusters (i.e. the number of specific network domains) [114]. This algorithm was developed by Dr. Stijn van Dongen [114].

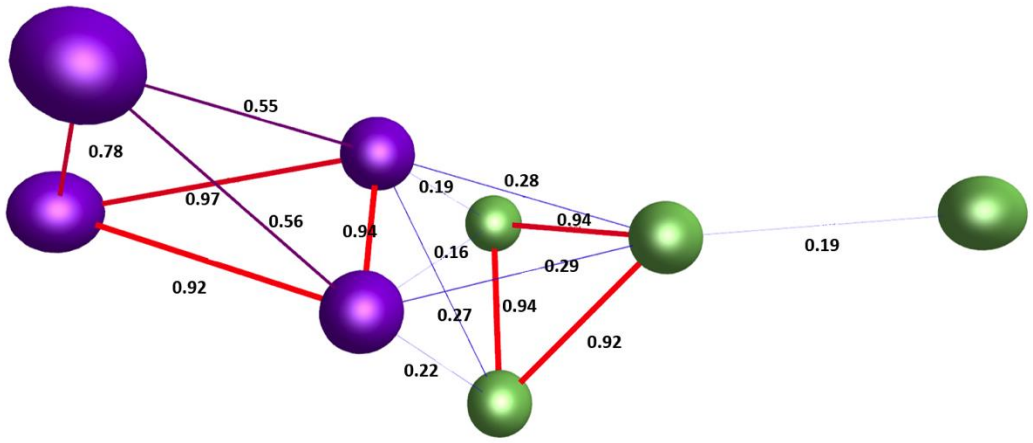


Figure 1.14: Example of network structure composed of nodes (genes) and edges (correlation between the genes). Heat colours between red (High similarity) and blue (Low similarity) demonstrate the relationship between each of the nodes. Two clusters (Purple and green nodes) were identified based on the MCL calculation of network structure. The Pearson correlation coefficient determines the strength of relationship between nodes.

The samples (e.g. nodes that represent the expression profile of genes in array data) are clustered mainly based on the properties of the network structure but not the similarity of gene expression profile. This means that the number and size of clusters are determined by 1) the connectivity between nodes and 2) the weight of each particular edge connecting two nodes defined by Pearson correlation (degree of similarities between two nodes) in a network structure. Clustered nodes (genes or transcripts) share a high similarity of their expression profile.

BioLayout *Express*^{3D} also enables the freedom to determine the granularity of clusters by adjusting the MCL-inflation value that changes the stringency of the cluster membership. Higher inflation value of MCL enhances the strength of correlation between clustered members dividing a network structure into smaller clusters. In contrast, low inflation value of MCL dilutes the strength of correlation between clustered members and result in larger size clusters.

RNAi and current approaches in cell-based genomic functional analyses

RNA interference (RNAi) cascades inhibit gene expression by targeting homologous mRNA molecules, triggering the 'knockdown' of target genes [115]. Historically, RNAi was first discovered where sense and antisense RNA were equally involved in the inhibition of gene expression in plant cells by Napoli *et al.*, 1990 [116] and *Caenorhabditis elegans* (*C. elegans*) by Guo *et al.*, 1995 [117]. In 1998, this phenomenon was first described as RNAi when the injection of double-stranded RNA (dsRNA) molecules into *C. elegans* induced the silencing of sequence specific genes [118]. RNAi that induced post-transcriptional gene silencing (PTGS) was also discovered in plants in 1999 [119]. In 2000, RNase III (Dicer) was identified to be involved in ATP-dependent dsRNA segregation, which resulted in dsDNA fragments of 21 to 23 nucleotide intervals in *Drosophila* embryos [120]. In 2001, RNAi was also found in mammalian cells [121]. In addition, Dicer was found to be evolutionarily conserved across different species e.g. mammals, plants, flies, and worms [122].

The RNA pathway generally consists of two stages, initiated by Dicer, which cleaves long dsRNA into 21 to 23 nucleotides of dsRNA molecules. After being unwound, each dsRNA molecule is disassociated into a pair of single-stranded RNAs (ssRNAs) i.e. the passenger and guide strands where only the guide strand is involved in the RNA induced silencing complex (RISC) pathway.

There are two types of small RNA molecules that play central roles in gene silencing i.e. short-hairpin RNAs (shRNAs) and small interfering RNA (siRNA). The siRNA is exogenous while shRNA is endogenous [123]. It is interesting that shRNAs are generated using mRNA pathways. However, shRNAs are not expressed as proteins. shRNAs are cleaved into shorter RNA molecules having a hairpin structure, also called shRNAs (Figure 1.9A) [124].

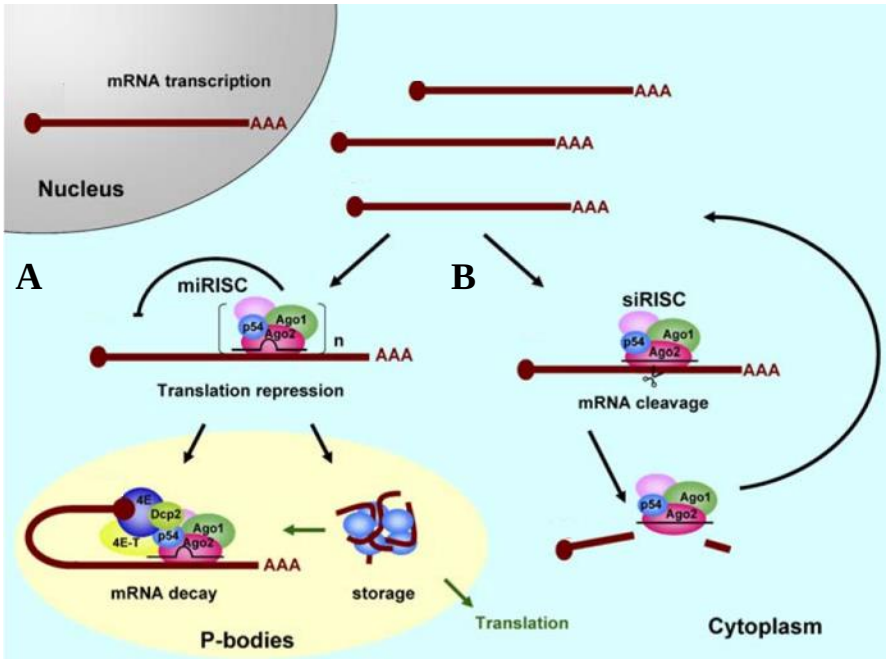


Figure 1.15: Schematic of human RISC pathways induced by siRNA and shRNA. A functional protein complex of RISC is composed of Ago2, Ago1, p54, proteins that have not been identified, and Dicer throughout the cytoplasm. **A)** The shRNA pathway represses the translational mRNA by storing the RISC/mRNA complexes in P-bodies or conducting the degradation. **B)** siRNA mediates the gene silencing by perfectly binding with homologous mRNA sequences in the cytoplasm. This figure is courtesy of Chu *et al.,* 2006 [123].

The difference between siRNA and shRNA pathways is that the shRNA conducts gene silencing by imperfectly pairing with the homologous mRNA (Figure 1.9A) [123]. To repress the target gene expression, multiple RISC complexes are involved in the shRNA pathway by binding to target mRNA, accumulating in P-bodies, where the RISC complexes form a hairpin shape structure in the middle of target mRNAs (Figure 1.15A). Cellular conditions determine whether repressed mRNAs are stored in P-bodies or degraded (Figure 1.15A). In contrast, siRNAs inhibit gene expression by perfectly pairing with homologous mRNA, leading to mRNA degradation in the cytoplasm (Figure 1.15B) [123].

RNAi approaches

To date, the genome-scale libraries of RNAi reagents (i.e. dsRNA molecules chemically synthesised) are built for high throughput system (HTS) approaches using genomic sequence data [125]. Benefitting from the miniaturisation and automation of technologies, RNAi screening can be used for functional genomic analyses in combination with HTS, identifying loss of function phenotypes [126]. Using the cell-based RNAi HTS, a large number of genes can be screened in small scale studies in parallel, which can be used to predict their putative

roles in biological events and processes e.g. RNAi that knockdowns the expression of a gene of interest affects the phenotype of cells [127].

The cell-based RNAi HTS can be applied in arrayed format as each gene is targeted by designed reagents in different wells where the cells are plated using microtiter plates e.g. 96 or 384 well plates [124,128,129]. The readout can be evaluated using colorimetric analysis to determine the intensity of fluorescence/luminescence at the total well level [130,131], and using imagine screening to determine the cellular or subcellular fluorescence [132,133]. As it is easy to categorise a large number of RNAi reagents used in arrayed format, large-scale RNAi screening can be performed e.g. to identify unknown cell cycle-associated genes in *Drosophila* [130,131], human cells [132-134], and *C. elegans* [135].

RNAi reagents

As more was learnt about the RNAi pathways that repress gene expression, the design of RNAi reagents and their libraries were evolved for various applications. In the 1990s, the long dsRNA segments were found to induce gene silencing in *C. elegans* [135] and Drosophila [130,131]. However, gene silencing by long dsRNA segments was not identified in mammalian cells as a non-sequence response was triggered by molecules greater than 30-base-pairs [136].

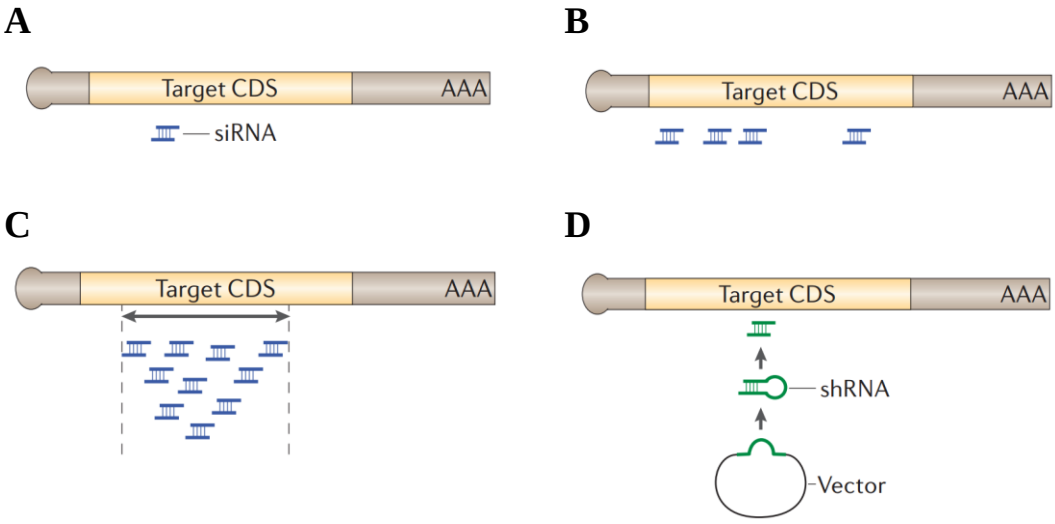


Figure 1.16: Mammalian cell-based RNAi screening reagents. A) The siRNA molecule consists of a canonical structure (19 bps long) where 2-bps overhang at the 3'end of both strands. **B)** siRNA is designed as a low-complexity pools of less than 10 molecules targeting the same transcripts. **C)** esiRNA can be used as high-complexity pools to target the same transcripts. **D)** shRNA is a vector-based library reagents which are introduced into cells virally. Diagrams are courtesy of Echeverri *et al.*, 2006 [125].

Long dsRNA molecules in the cytoplasm in mammalian cells appear to activate various pathways of which repress protein synthesis, collectively known as the interferon response [137]. To provide a solution, short dsRNA were designed, inspired by the siRNA pathway in plant cells [119]. Therefore, short dsRNA segments are commonly used in mammalian cell-based research [132-134].

In most mammalian cell RNAi screening, short dsRNAs are designed following the principle that the dsRNA sequences closely mimic 21 nucleotides long endogenous sequences of homologous mRNA molecules with 2-bases extended at both 3' ends (Figure 1.16A) [121]. There are the 4 popular types of short dsRNAs reagents i.e. standard siRNAs, low-complexity siRNA pools, and endoribonuclease-prepared siRNAs (esiRNAs), shRNAs (Figure 1.16).

Delivery system for RNAi reagents

In addition to the design of the RNAi reagents, the delivery system for dsRNA molecules is also important in cell-based RNAi screening. Appropriate delivery systems determine their transfection/transduction efficiency. There are two common approaches including lipid-conducted transfection and viral-conducted transduction [126]. For example, siRNAs and esiRNAs can be delivered by lipid-conducted transfection [127,132]. shRNAs can be introduced into the cells by viral-conducted transduction [133,134].

Endoribonuclease-prepared siRNA

RNAi induced gene silencing is a well-known mechanism that mediates mRNA degradation in plants [119,138] or animals [121,139] and is initiated by double-stranded RNA homologous with a targeted gene [121].

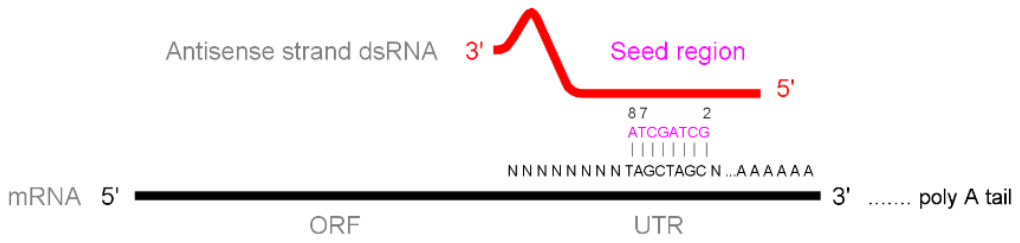


Figure 1.17: siRNA building to seed region on off-targeted of targeted mRNA. Seed region on the antisense strand of siRNA (nucleotides between 2nd and 7th or 8th, shown in pink N) that is highly complementary to the sequences on the 3' untranslated region of off-targeted mRNA.

However, genes sharing similar sequence to a targeted gene may result in the off-target effects [140]. Microarray profiling of gene silencing experiments has demonstrated numerous off-target transcripts with a reduced expression level [115,141,142]. Other reports suggest that excluding the near-perfect matches between targeted sequences on mRNA and antisense of siRNA, off-targeting is mediated due to a perfect match between the antisense strand of siRNA, mediated by the 2^{nd} ~8^{th} bps (the seed region) and the 3' UTR (Untranslated region) of the mRNA (Figure 1.17) [143].

To minimize off-target effects, Ralf *et al.,* 2007 [132] and Frank *et al.,* 2006 [144] developed esiRNA. This involves enzymatically-preparing dsRNA to provide a heterogeneous population of siRNAs that target different regions of an mRNA, and results in a dilution of off-target effects. esiRNA has been shown to efficiently knock down the expression of targeted genes in human cells [145]. Since gene silencing was first demonstrated in mammalian cells by Elbashir *et al.,* 2001 [121], this technology has been widely used to investigate the function of genes in biological systems combined with HTS [125].

Scoring cut-off used to determine the RNAi effect on cell proliferation

The application of RNAi screening combined with HTS has been used to investigate gene function in cells in many different species e.g. *Drosophila* [130,139], *C. elegans* [131], mouse [133], and human [132,146]. In order to deal with a large number of measurements in HTS, most researchers use a Z-score to determine the hits at a particular threshold from the screen [130-135,147,148]. A Z-score ($Z = ((x-\mu))/\sigma$; *x* represents sample, μ represents the mean of population of samples, σ represents the standard deviation of population of samples) is a statistical measurement of a score's relationship to the mean in a group of scores. A Z-score can be positive or negative, indicating whether it is above or below the mean and by how many standard deviations [149,150]. For example, if the Z-score is +3, then this means that the score is three times the standard deviation above the score of the population mean.

Z-scores are often calculated for a specific time point. However, end-point analysis can only provide a "snapshot" at a particular time point and may miss differences that might be present at other time during the experiment. In order to analyse the real time data from the real time cell analysis (RTCA) experiments, it is therefore necessary to calculate a score based upon all or part of RTCA profiles.

There have been a number of high throughput RNAi screening experiments monitored over a time series to overcome the drawback of endpoint analysis. For example Beate *et al.,* 2010

[151] used RNAi screening combined with automatic time-lapse microscopy and computational image processing to profile the phenotype of HeLa cells at a number of time points. In this experiment, around 21,000 human protein-coding genes were knocked down and images were recorded of chromosomes tagged with green fluorescence protein (GFP) on core histone 2B to determine the stages of the cell cycle. As histone duplication is coupled with DNA replication, changes to cell cycle profile were used to analyse the effects of knocking down individual genes on proliferation (chromosomes 2N: G_0, G_1; 4N: S, G_2, M) [151,152]. In another RNAi screening experiment, Hutchins *et al.*, 2010 used time-lapse microscopy to profile the localisation of proteins tagged with GFP in HeLa cells. The strength of GFP signals profiled by immune fluorescent microscopy (IFM) over a time series, were scored and used to determine the role of labelled proteins in association with chromosome segregation [153]. However, to find the best time-point for high-content bio-image and the length transfection in RNAi screening is still a challenge of experiment design [115].

RTCA system

The best way to analyse the effect of RNAi on the cell cycle is therefore over a time series. The RTCA is a technology that indirectly monitors cell proliferation based on changes in electrical impedance as cells divide in a culture dish. Electrical impedance of cells is detected via a gold microelectrode array that covers the base of well and records the cell impedance index (CI) based on the electrical conductance of lipid bilayer of membrane of living cells. [145,154-157]. In an RTCA experiment, CI increases as cells grow; this is roughly equivalent to cell proliferation although CI also increases if cells spread out and adhere more tightly to base of well (e.g. macrophages). We have used RTCA to analyse the effect of knocking down specific proteins.

Since RTCA is a highly sensitive system, minimise data normalization is required to be essential to minimalize the difference between wells and plates [115]. The trace of CI at the point of normalization is readjusted from recorded value to 1.

Determining the hits from RNAi screening

However, it has been found that standard deviation increases in replicates as time progresses from the normalization point. To overcome this, Zhang, *et al.*, 2011 suggest using the cell impedance index growth rate (CIGR) instead of CI [115]. The CIGR represents the rate of change of CI over time series, reflecting the kinetic of cell proliferation before and after gene silencing. It was demonstrated by Zhang, *et al.*, 2011 that the standard deviation of CIGR did

not steadily accumulate following a fixed time point chosen for normalization but rather stayed constant [115]. This makes it possible to normalize at a fixed time point that does not result in an increase in standard deviation at later time points.

Due to the limitation that Z-score can only determine which genes in the RNAi screening are hits by a single measurement value at a fixed time point, Zhang, *et al.*, 2011 suggested calculating the average of CIGR over time series as a single value. Therefore averaged CIGR can be used to analyse effect of gene silencing on the cell proliferation [115].

Subcellular protein localisation

There are numerous biological processes happening in parallel to drive cell proliferation and other essential housekeeping events. The cell is a space that consists of different organelles with many proteins localised to specific locations. In general, proteins involved in the same biological processes, are also co-localised e.g. proteins localised in the centrosomes are generally involved in chromosome segregation (see Chapter 5).

Subcellular localisation suggests a fundamental feature of a protein. To investigate an unknown protein and its function in a relevant biological processes, it is useful to localise their position in the cells for further experimental design.

Protein subcellular localisation has become one of the most popular techniques, characterising the function of unknown proteins. There are various recombinant report proteins with a fluorescent signal, e.g. GFP that can be used. GFP was firstly isolated from the jellyfish *Aequorea Victoria* [158] and is widely used as a fluorescent tag to localise proteins of interest plants [159,160], yeast [161], and mammalian cells [162,163].

GFP is commonly co-expressed with proteins *in vivo* by tagging, and localising by microscopy inspection. GFP has a self-emission fluorescent chromophore due to the specific sequence of three amino acids i.e. serine, tyrosine, and glycine [164]. As it is weakly phototoxic, GFP is commonly used to localise the protein of interest in living cells [165]. In addition, the self-emission signal does not require additional stimulators or cofactors to be activated [164].

To avoid the interruption of the localisation signal (i.e. GFP) and over expression of protein that may saturate the cellular transportation machinery, a method was developed for tagging GFP at the carboxyl terminal end, positioning the GFP behind the endogenous promoter

[166]. In addition, the coding sequence of GFP was inserted in-frame just in front of the stop codon in open reading frames (ORFs) [166].

With a robust expression system, protein subcellular localisation has been applied at a genome-wide level [167]. In doing so, a large-scale project 'ORFeome' was carried out that clone the entire set of ORFs, expressed in fission yeast [167]. 'ORFeome' cloning is a resource developed for the construction and sharing of sets of clones that can be used to study proteins of interest in a genome-wide manner [168]. The open-reading frame (ORF) is different to the protein-coding sequence (CDS). The CDS covers a region of sequence between the translation start and termination site. However, ORF includes the contiguous region but without the stop code. This is because the aim of ORFeome cloning is to express clones with the fusion protein of interest [169]. It is very important that the set of ORF clones can direct the synthesis of fused proteins. In order to express the clone with fused proteins i.e. GFP, the ORF clone must carry ORF without the original stop codon [169]. Therefore the expression of the GFP signal can be used for subcellular protein localisation.

To perform the ORFeome cloning, it is important to choose an effective high-throughput cloning system as an appropriate expression vector e.g. Gateway cloning system. The human ORFeome project version 8.1 is a public resource, containing 12,230 clonal isolates (i.e. clonally-derived and sequence-verified ORFs cloned as a set of Gateway Entry clones) representing at least 11,149 human genes [170].

The Gateway cloning System

A traditional molecular cloning experiment includes four elements: 1) fragmentation of DNA, 2) ligation of fragmented DNA with a designed sequence, 3) transfection of recombinant DNA molecules into host cells, and 4) selection of host cells which are successfully transfected with the recombinant DNA molecules [171].

One of many molecule-cloning methods is Gateway cloning technology. This technology allows the functional analysis of gene and protein expression. Gateway cloning evolved from typical molecular cloning technologies and was developed based on the concept of site-specific recombination used by bacteriophage λ to infect the bacteria [172].

Gateway cloning has a number of advantages enhancing the accuracy and robustness in generating recombinant DNA molecules. Firstly, this method provides a fast and efficient solution to deliver DNA sequences into multiple vector systems. Secondly, the cloning system is highly compatible with various types of DNA sequences, e.g. PCR products and

cDNA clones. Thirdly, transfection of DNA sequence, into destination vectors is relatively easy. Finally, it is suitable for HTS analysis. Given these advantages, Gateway cloning is widely used for subcellular protein localisation for functional analysis [173].

Pathway modelling

One of the many challenges encountered when studying an entire biological system is finding a solution to collating the huge amount of information obtained from textual descriptions and experimental observations. Pathway modelling enables the visualization of all the biological discoveries across different studies. The problem is that the complexity of biological events occurring during cell proliferation can be depicted in various network structures.

It is known that using a pathway can help utilize the information available following previous experimental discoveries. Pathways still rely on the extensive textual description to explain aspects that cannot simply be demonstrated within the pathway itself. In addition, the design of the pathway is strongly influenced by the need of the developers and this increases the difficulty of information accessibility for other biologists. A pathway map should aim to provide a platform where biologists can utilize it to share their biological research and further research in the field.

Figure 1.18 shows various pathway diagrams of the centrosome life cycle depicted according to different sources. Focusing on different biological events within the centrosome life cycle, it is difficult to understand a particular event and its relationship in association with other biological processes. In order to address this issue, a new approach to pathway notation inspired by the layout of an electronic circuit from the electronic industry was developed [174-176].

In a consensus view, a pathway map is a regulatory network composed of proteins and genes connected by edges representing biological events [177,178]. The proteins are defined as components and interactions between them are depicted as arrows. To identify these components in biological events, the first task is to define interactions within these protein-protein interaction networks [177,178]. Since the concept of pathway mapping was introduced, more and more biologists have started to develop approaches to depict the regulatory network of biological processes. To this end, many types of pathway map schemes have been developed to serve the biologist providing a large-scale regulatory

network map of events e.g. systems biology graphical notation (SBGN) [179], and modified Edinburgh Pathway Notation (mEPN) [177].

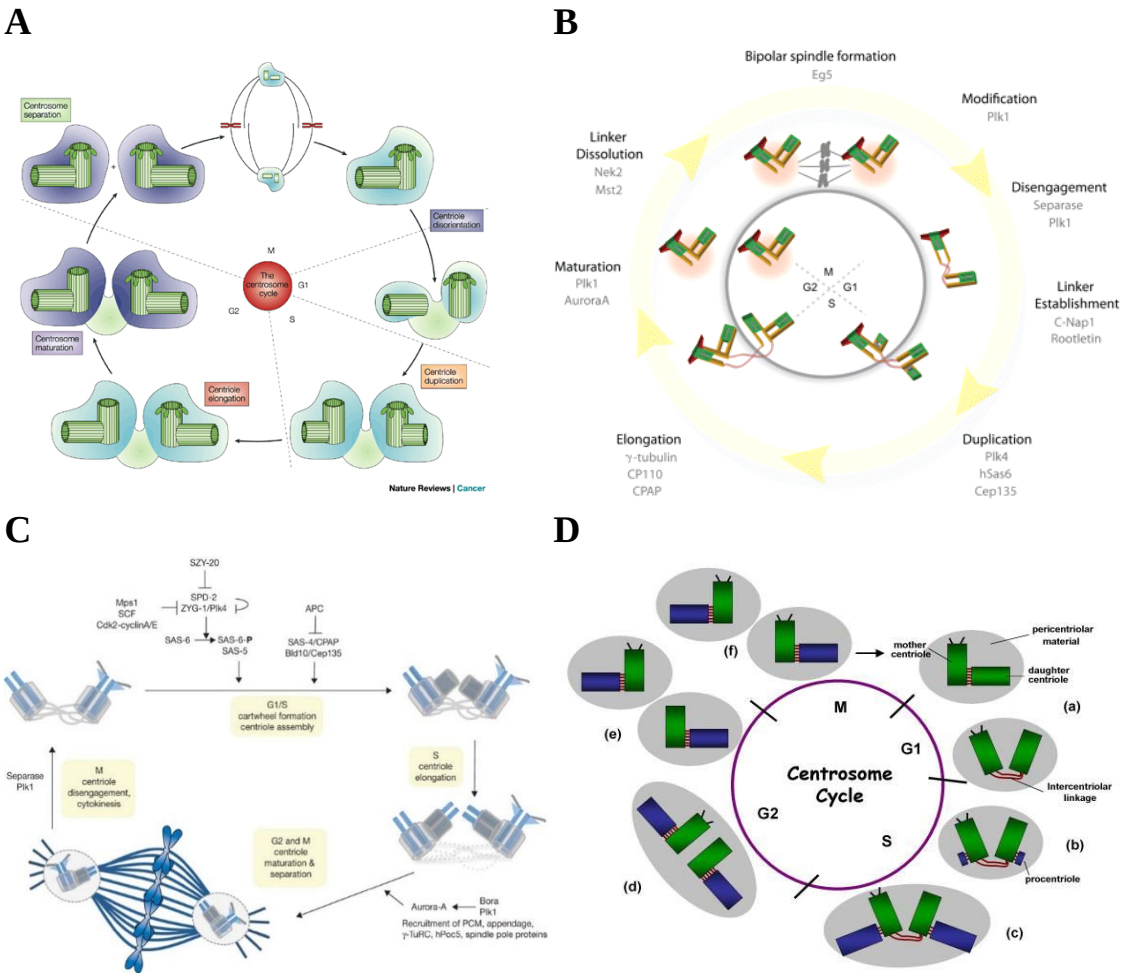


Figure 1.18: Centrosome life cycle demonstrated in different formats. The 'same' pathway map shares poor consistence at layout consistency, standard graphical notation, and notation system (if any used) for pathway components. **A)** Centrosome duplication life cycle extracted from articles; Nigg EA *et, al.,* 2002 [47]. **B)** Mardin BR *et, al.,* 2012 [180], **C)** Jaspersen SL *et, al.,* 2008 [181], and **D)** from Crasta K et, al., 2006 [182].

The SBGN is a graphical representation language developed by a community of computer scientists, modellers, and biochemists aimed to visualise the biological processes as a pathway map [179]. The mEPN is an alternative graphical language based on a pathway notation system first proposed by Moodie *et al.,* [176] in 2006 and first published in this form by Raza *et al.,* 2008 [174]. Both schemes evolved from similar concepts, such as components, processes, and cellular compartments, which are used to construct pathway nodes.

To date, there are more than 60 pathway diagram databases accessible online [183], e.g. WikiPathways [184], REACTOME [185], and Kyoto Encyclopaedia of Genes and Genomes (KEGG) [186]. WikiPathways is a collective resources of biological pathways depicted by various participants in a community [184,187]. The publishing, updating, and maintaining of the pathway maps are reviewed by a group of experienced users [187]. REACTOME is an online database that provides pathway maps depending on different organisms [185]. The pathway maps are based on human biology [188]. Prior to the release of pathway maps which are reviewed by experienced pathway curators with PhD level training. KEGG is one of the most popular pathway databases and was developed since 1995. This database includes pathway maps of cellular processes e.g. metabolism, membrane transport, and the cell cycle [189]. The pathway maps were developed by a group of pathway biologists.

Although pathway schemes are well developed in most cases, there is no standard international graphical notation system [183]. They are either depicted centrally, such as KEGG or by the community e.g. WikiPathways. This increases the threshold of information accessibility between different pathway databases due to their incompatibility between each other. In addition, regular pathway maps cannot be used for computational simulation as high performance computers are able for such approaches forecasting the regulation behaviour of a set of proteins in the pathways [190-192].

To this end, the mEPN pathway scheme was developed to provide pathway maps, which can be used to simulate the regulation behaviour of protein-protein interactions in BioLayout *Express*^{3D} [113,177,179,193]. In addition, BioLayout *Express*^{3D} can convert the mEPN pathway to other popular pathway schemes e.g. SBGN. An advantage of the mEPN pathway map is that it can be used for multi-task applications (e.g. simply mapping the biological processes of interest and storing the information in maps by attaching them in the nodes or edges), the mEPN scheme was used to incorporate and interpret the results i.e. array analyses (Chapter 2), RNAi screening (Chapter 3), and protein subcellular localisation (Chapter 4). In chapter 5 a large-scale mEPN pathway map was developed to summarise the entire story of this thesis.

Aims and objectives of this works

Transcriptomics analyses by the group commonly observed a cluster of genes in many datasets that were highly enriched in cell cycle genes. Amongst this cluster of known cell cycle genes we often observed genes of unknown function. This thesis is based on the hypothesis that there are as yet unidentified human genes that play a role in cell cycle progression and that they will be co-expressed with known cell cycle genes. In order explore this hypothesis we have:

- Surveyed the human cell cycle regulated transcriptional network using latest array technology and use more advanced data analysis techniques.
- Characterised the putative novel cell cycle-associated genes particularly those associated with S- and G_2/M- phase.
- Identified the role of putative novel genes using RNA interference monitored by RTCA systems, flow cytometry, and RT-qPCR to check the expression level of targeted genes.
- Characterised the cellular localisation of novel cell cycle-associated proteins.
- Developed a system-wide graphical model of the centrosome regulation involved in the cell proliferation.

Chapter 2.

Transcriptional network analysis of cell cycle-associated genes expressed in a synchronized major population of human cells

Introduction

A scheduled progression of the cell cycle across its different phases is essential to the growth and development of all organisms [194]. Many of these processes have been well studied at the molecular level, such as quiescent program regulation [22], DNA replication [195], chromosome segregation [196], and the checkpoint systems [197]. The cell cycle is one of the best-studied pathways by transcriptomic analyses with hundreds of genes up regulated as cells pass from S-phase to M-phase.

Studying cell cycle-regulation at a genome-wide transcriptional level has been performed in human cells by microarray e.g. HeLa cells [94] and human fibroblasts [22,96,98]. To study cell cycle-regulated transcription, cells must be synchronized before cell cycle-associated genes can be identified [7]. Therefore the synchrony of cell cycle is important in array analysis.

There are two factors that determine the synchrony of mammalian cell populations. Firstly, normal diploid mammalian cells lose their synchronization relatively soon after re-entering the cell cycle following cell cycle arrest [96]. Furthermore, it is thought that the degree of synchrony after re-entering the cell cycle varies by cell type [89]. In addition, when a cell divides it has been shown that the two daughter cells may have different cell kinetics [89]. Secondly, it is known that only 50 to 70% of mammalian cells re-enter the cell cycle after cell cycle arrest [98]. For mammalian cells, chemical inhibitors can be used to mediate partial cell cycle synchronization. For example, metabolic inhibitors thymidine [198] and aphidicolin [199] both arrest the cells at G_1/S-border. In addition, nocodazole can be used to arrest the cells at M-phase [200]. However, other methods utilize a non-chemical inhibitor to mediate partial synchronization. Mechanical synchronization methods such as mitotic shake-off can select cells at G_2/M-border but yield a low number of cells [201]. Fibroblast cells are arrested in G_0/G_1-phase by serum deprivation but re-enter the cell cycle when fed serum. The response of fibroblasts to serum stimulation allows the study of cell cycle-regulated gene expression after the cells are released from arrest [96].

In previous studies, around 700 cell cycle-regulated genes relevant to G_1-, S-, G_2-, and M-phase were identified by Cho RJ, *et al.*, 2001 in human fibroblasts synchronized by a double thymidine block [202]. In addition, due to the limitation of having an incomplete human gene expression data set (U133A arrays), the array that Bar-Joseph, *et al.*, 2008 used only covered two thirds of total gene expression. Hence, in this study, a complete human gene data set can render us a more complete view of cell cycle-associated gene expression.

The aim of this chapter is to survey the human cell cycle regulated transcriptional network using latest array technology and use more advanced data analysis techniques. To create a list of all the known cell cycle-associated proteins particularly those associated with S-, G_2-, and M-phase.

In order to identify the best cell lines for array analyses, serum starvation is tested to find the best cells for synchronization. We have tested human embryonic kidney 293T (HEK293T), human acute monocytic leukemia cell line (THP-1) and neonatal human dermal fibroblast (NHDF). Conditions were optimised to try to create a synchronous major population of cells progressing through the different phases of the cell cycle. In addition, 40 new microarrays for two time course experiments were performed, at 6 h intervals (24 arrays) and at 2 h intervals (16 arrays).

Materials and Methods

Overview of approach to identify cell cycle associated genes

The flow chart below shows an overview of the strategy to identify cell cycle-associated genes and is composed of a number of stages (Figure 2.1). Firstly, to identify the most appropriate cell whose mitotic division could be arrested by serum-deprivation and then reactivated by serum refeeding to enter the cell cycle. Secondly these cells would be used to assess cell cycle-associated transcription by microarray analysis and then mined to derive a list of cell-associated genes.

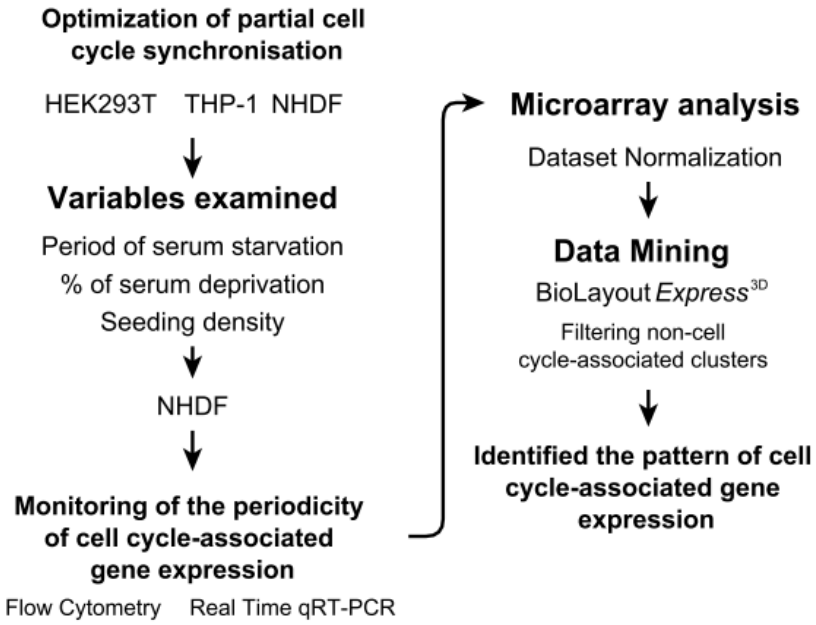


Figure 2.1: Overview of experiments toward, the identification of cell cycle-associated genes. HEK293T, THP-1, and NHDF cells were selected to monitor synchrony after re-entering the cell cycle. Conditions of synchronization included the %, length of serum starvation, and seeding density were optimised. NHDF cells were found to be able to generate a major population with synchronised cells. To identify cell cycle associated gene expression, two time course microarray experiments were performed and data mining was used to monitor the pattern of gene expression. Normalized dataset from microarrays was surveyed and analysed in order to find novel cell cycle-associated genes.

Cell Culture

HEK293T cell culture and cell cycle synchronization

HEK293T cells are the HEK293 cell line containing the simian vacuolating virus 40 (SV40) large t-antigen [203]. HEK293T cells between passage number 7 to 12 were cultured in DMEM (Dulbecco's Modified Eagles Medium, Sigma-Aldrich, Gillingham, UK), including 10% (v/v) Fetal Bovine Serum (FBS) (GE Healthcare, Little Chalfont, UK), antibiotics (25 U/ml penicillin and 25 µg/ml streptomycin) (Life technologies, Paisley, UK), 0.1 mM Non-Essential Amino Acids (NEAA, Glycine 0.1 mM, L-Alanine 0.1 mM, L-Asparagine 0.1 mM, L-Aspartic acid 0.1 mM, L-Glutamic Acid 0.1 mM, L-Proline 0.1 mM, L-Serine 0.1 mM) (Life technologies), 1 mM glutamax (Life technologies), and fungizone antimycotic (1.25 µg/ml) (Life technologies). In the study of HEK293T cells, different percentages of serum were used to starve the cells. Cells were cultured until the proportion of coverage was 30~40%. Serum starvation was performed by washing with serum free medium twice and culturing in medium with 0.5% serum for 48 h. The starved cells were subsequently placed in medium with 10% serum (refeeding) and cultured for 20 h before harvesting.

THP-1 cell culture and cell cycle synchronization

The THP-1 cell line was originally isolated by Tsuchiya *et al.*, 1980, from a patient with acute monocytic leukemia [9]. For culture and passage, THP-1 cells at passage number between 7 and 12 were cultured in 175cm^2 Thermo tissue culture flasks at a density of 1x10^5 cells/ml. The cells were cultured in RPMI 1640 medium (Sigma-Aldrich), with 10% (v/v) FBS (GE Healthcare), 1 mM glutaxmax (Life technologies), and antibiotics (25 U/ml penicillin and 25 µg/ml streptomycin) (Life technologies). Cells were cultured between a density of 1x10^5 and 1x10^6 cells/ml. Medium (30 ml) with 0.2% (v/v) FBS was used to starve the cell for 48 h followed by refeeding with serum. THP-1 cells were harvested at 0, 6, and 24 h intervals.

NHDF cell culture and cell cycle synchronization

NHDF cells isolated from neonate foreskins were kindly provided by Dr. Finn Grey's lab (The Roslin Institute, Edinburgh, Scotland, UK) [204]. For culture and passage, adherent cells were plated on 175cm^2 Thermo tissue culture flasks at density of 6 x 10^3 cells/cm^2. NHDF cells at passage number between 7 and 12 were cultured in DMEM (Sigma-Aldrich) with 10% (v/v) FBS (GE Healthcare), and antibiotics (25 U/ml penicillin and 25 µg/ml streptomycin) (Life technologies). Cells within the flasks were grown under the condition of 5% CO_2, 95% air at 37^{o}C. Cells were cultured until 30~40% confluent before placing in

0.5% (v/v) serum for 48 h starvation and then re-fed by placing them in complete medium and harvesting at different time points.

Flow cytometry

Flow cytometry was used to monitor and determine the number of cells in each phase of the cell cycle at each time point showing the effect of serum starvation and refeeding on NHDF, HEK293T and THP-1 cells. Cells were re-suspended in 50 µl PBS and then mixed with 1 ml 70% (v/v) cold ethanol (EtOH). Samples were fixed at 4°C for at least 24 h. Cells were centrifuged at 400xg for 5 min and the supernatant discarded. The cell pellet was then re-suspended in 100 µl 1X PBS. After pipetting the mixture vigorously to break up the cells, they were washed twice in 1 ml of PBS, and re-suspended in 100 µl of staining solution (38 mM sodium citrate buffer containing 138 µM propidium iodide, 10 µg/ ml RNaseA solution, and pH 8.0). After adding a further 200 µl staining solution, samples were placed in the dark at room temperature for at least 2 h. Synchrony of the cell cycle progression was analysed using BD FACS Aria lllμ Cell Sorter (BD bioscience, Oxford), and data was analysed by BD FACSDiVa^{TM} software (BD bioscience).

Microarray Analysis

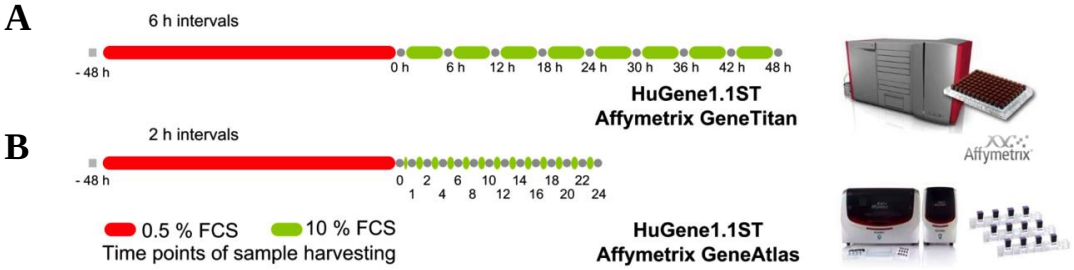


Figure 2.2: Design of microarray studies. **A)** Samples of NHDF were harvested at 6 h intervals following serum-refeeding (10% FCS) of cells cultured in 0.5% serum for 48 h. NHDF cells were analysed by microarray (Affymetrix® Human Gene ST 1.1 Arrays) on Affymetrix GeneTitan®system in the facility of ARK-Genomics at the Roslin Institute. **B)** A second microarray was performed where cells were treated as above, but samples were collected every 2 h after serum refeeding and analysed on Affymetrix® Human Gene ST 1.1 Arrays on the Affymetrix GeneAtlas® microarray system. Arrays were run as part of the Wellcome Trust Advanced Course on Functional Genomics and System Biology, Hinxton, Cambridge. Figures of Affymetrix GeneTitan/Atlas system are courtesy of Affymetrix, Inc. [205,206].

NHDF samples were harvested at different time points following serum refeeding, snap frozen in dry ice, and stored at -80°C until use (Figure 2.2). Total RNA was isolated using the RNeasy Mini Kit (QIAGEN, Manchester, UK) according to manufacturer's instructions. Quantification of total RNA yield was performed using the ND-1000 Spectrohotometer

(Nano-Drop Technologies, Wilmington, North Carolina, USA) based on the ratio of sample absorbance at 260 nm. The ratio of absorbance at 260 and 280 nm was used to assess the purity of RNA.

RNA quality was assessed using the Agilent RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, California, USA) and ran on the Agilent 2100 Bioanalyser using electrophoretic separation on micro-fabricated chips. The quality of RNA samples was determined by the ratio of 18S and 28S ribosomal subunits. Labelled cDNA fragments for microarray were generated by the reverse transcription of total RNA (500 ng) using the Ambion® WT Expression Kit (Life technologies), fragmentation, and then labelled by TdT DNA labelling reagent using GeneChip® WT Terminal Labelling Kit (Affymetrix, Bucking hamshire, UK) according to manufacturer's instruction (Affymetrix, GeneChip® WT Terminal Labelling and Hybridization User Manual).

mRNA expression analysis by quantitative RT-qPCR

RNeasy Mini Kit (QIAGEN) was used to isolate total RNA from a major population of synchronized NHDF cells following time points after serum refeeding. Two-step RT-PCR was performed to generate cDNA. For RNA samples and given primers, each reaction was performed in 30 µl volumes using MicroAmp Optical Caps (Applied Biosystems).

0.5 µg of total RNA mixed with ddH_2O in total volume 18.9 µl and was initially heated to 72^{o}C for 5 min. It was then incubated for 2 minutes on ice. For subsequent reverse transcription, each reaction used primer Ploy (dT) (5.0 µM) 1.5 µl, RNAse inhibitor 0.6 µl (Promega) and dNTP 1 µl. 2 µl of the Reverse Transcriptase MMLV (13.3 u/µl, Invitrogen) was added in 1x Reaction/PCR buffer (Invitrogen) for each reaction. Samples were incubated at 42^{o}C for 65 min and 95 ^{o}C for 5 min, and stored storing at -20 ^{o}C until further use.

Using the primers shown in Table 2.1, real time reverse transcription qPCR was performed using SYBR® Green I (Roche) and the LightCycler® 480 II real-time PCR system (Roche, Hertfordshire, UK). In each PCR, 2 µl of cDNA was combined with gene-specific primers (0.5 µM) and 12.5 µl of QuantiTect SYBR Green PCR Master Mix to a total volume of 25 µl.

To enable calculation of efficiency of the PCR reaction and to check that the PCR reaction was quantitative, a representative cDNA was serially diluted. Serial dilution was performed cDNA from the time point predicted to provide the maximum expression level of the gene

product being amplified. Other controls performed in each run included water blank and samples which did not have Reverse Transcriptase.

The conditions used to perform qPCR were as follows: First, denaturation step 95°C for 5 min; second, annealing step 59°C for 10 sec; third, extension step 72°C for 20 sec; and fourth back to first step for 45 cycles with 10 sec for the 95°C denaturation step. At the end of each run, melting curve analysis was performed between 65°C and 97°C. LightCycler® 480 (Roche, Hertfordshire, UK) analysis software version 1.5, was used to compare amplification in experimental samples during the log-linear phase to the standard curve from the dilution series of control cDNA. Comparisons were displayed as bar charts with standard deviation. Comparison between the expression level of amplified genes and *GAPDH* (Glyceraldehyde-3-phosphate dehydrogenase) as a ratio were displayed as histograms with standards deviation.

Marker	Sequence (5' to 3')
CDK1	U-TATGCCTTGGTCAGAGTAATAACT D-ATACGGAACCAGTCTCATTATTGA
BUB1	U-TCCTCCTTTGGCAAATGCTATT D-AGGAGGAAACCGTTTACGATAA
GAPDH	Hs_GAPDH_2_SG QuantiTect Primer Assay (200) (QIAGEN, Manchester, UK)

Table 2.1: Sequence of primers for qPCR. U and D refer to upstream and downstream primers. qPCR, quantitative polymerase chain reaction.

Quality assessment and data processing

The Bioconductor package ‘oligo’ was used to perform the quality assessment of array dataset [207]. To determine the quality of data, pseudo image, density plots, and boxplots of the probe reading intensities of arrays were performed using R software, an environment for statistical computing and graphics plotting graphics (version 2.14.2) [208]. Following QC, the datasets including the array samples derived from NHDF cells at 2 or 6 h intervals were combined and normalized together utilizing robust multi-array average (RMA) [209]. Array data were saved as '.expression' file, and then investigated by network analysis using BioLayout *Express*^{3D} [112,113]. Based on the similarity of gene expression, the expression profile of genes correlated with a Pearson number above or equal to 0.7 ($r \geq 0.7$) were included and saved in the format as '.pearson' file. Cell cycle-associated genes were identified in a number of clusters.

Using network analysis to identify S-, G_2-, and M-phase associated genes

Probes of the RMA normalized dataset (total 33,298 transcripts) with signal intensity read levels below 35 were firstly removed from all array samples (11,456 transcripts removed) and then dataset clustered using BioLayout *Express*^{3D} with settings of Pearson correlation (r) = 0.85 and MCL-inflation value = 3.0 (Markov Clustering Algorithm). The size of clusters determines the ranking of clusters i.e. the cluster with largest number of nodes is labelled as cluster 001. Graphs of each dataset were explored extensively in order to identify the members of the gene clusters and their relevance to the cell cycle progression.

Clusters of genes with inconsistent expression profiles between the two data sets were removed (4,400 transcripts). After removal of negative and positive controls of transcripts providing reference signal for array scanning machine from the dataset (2,620 transcripts), remaining transcripts (14,822 transcripts) were clustered at lower constraint with MCL-inflation value = 1.7 in order to include most of the genes correlated with the periodicity of known cell cycle regulatory genes. Cluster associated with particular phase of the cell cycle was identified by presence of known cell cycle genes within the cluster (see List of additional information [22,94]), and then coloured blue (G_0), red (G_1-Early), purple (G_1-late), and green (S, G_2, and M) (Figure 2.16). We specifically emphasised on the investigation of genes that up-regulated during S-, G_2-, and M-phase.

To summarise the major biological events which clustered genes were involved in, clustered genes were manually annotated as unrelated, unknown, putative association, or known S-, G_2-, and M-phase associated genes, according to GeneCards (http://www.genecards.org/) [210], universal protein resource (UniProt, http://www.uniprot.org/) [211], REACTOME (http://www.reactome.org/) [185], database for annotation, visualization and integrated discovery (DAVID, http://david.abcc.ncifcrf.gov/) [212], and gene ontology (GO, http://www.geneontology.org/) [213].

Using network structure analysis to re-investigate published array data

To further investigate the effects of chemically induced cell cycle synchronisation at the transcriptional level in fibroblast cells against serum deprivation, the array dataset from Bar-Joseph, *et al.*, 2008 was downloaded, normalised, and annotated [98]. Raw data were normalized by RMA normalization resulting in 22,277 transcripts. Extracted transcripts were then associated with gene names (19,940 transcripts, 12493 unique genes). This was saved as an '.expression' file. Array samples with their maximum expression levels below 35 were filtered out. Dataset saved as '.expression' format were clustered at the condition of r = 0.85, MCL-inflation value = 1.6.

Functional annotation of clustered cell cycle-associated genes by GO slim mapping and regular GO enrichment analysis

Following the approaches that identified four major cell cycle-associated clusters based on the temporal expression profile and clustered key cell cycle-regulated genes, a simplified version of GO was firstly used to provide a broad review of the ontology content without the detail of specific terms [214]. GO slims are composed of a subset of the terms compared to the entire GO [214].

GO slims are useful to summarize the results of GO annotation of a microarray analysis when various classifications of gene product functions are needed. In order to further profile the genes up-regulated in the progression of the cell cycle, the list of genes was mapped to functional categories of biological process defined in GO slims.

Classification of clustered genes according to manual annotation

To further annotate the function of genes clustered with cell cycle-association, genes of interest that were co-expressed during the S-, G_2-, and M-phase were manually curated based on reference mining and other online resources.

It is important to define the rule by which genes were categorised according to data curation e.g. reference mining. Genes that have known function relevant to non-cell cycle associated events, but expression periodicity in our data implying that they are associated with mitotic activity were classified as 'unrelated function'.

Genes that with putative association (i.e. indirect evidence from experimental observation in publications suggest that they are associated with the cell cycle events) to the cell cycle

according to publications were labelled as 'putative association'. Genes that were not previously characterised but expression periodicity suggested they were in association with the cell cycle were classified as 'unknown function'. Genes, which had been completely characterised as cell cycle-regulated elements, were classified as 'known S-, G_2-, and M-phase association'.

To further investigate and refine the list clustered genes, Prof. Tom Freeman's group filtered the list by comparing with datasets of other human cells and tissues, the list of genes in the S/G_2/M cluster was refined; genes were only kept in the list if they were found to be co-expressed in cells and tissues with high levels of cell proliferation [215]. This work included the participation of all the team members in Prof. Tom Freeman's group for over two years to polish and refine the list of genes identified in two array data sets.

Evidence from the literature (e.g. PubMed) and other online resources (e.g. REACTOME [188]) were carefully reviewed by members of the Freeman group, where each member shared an equal number of genes listed in this thesis.

Information on gene, were categorised into columns according to their features e.g. pathways proteins were involved in cellular locations where expressed proteins were trafficked, protein complexes where expressed proteins were recruited, cell cycle phase where they were up-regulated, chromosome location of genes, and the accession number of each gene (e.g. ENSEMBL gene ID [216]).

From the completed table, three of each gene's major functions associated with biological events were selected and annotated. This work was manually reviewed by each member for this project.

Results

In this experiment, a number of cells lines were selected for analyses. HEK293T, THP-1, and NHDF cells were examined for the effect of variables on the synchrony of cells after starvation such as period of serum starvation and the percentage of serum used in deprivation. To identify cell cycle-associated gene expression, two time course microarray experiments were performed and data mining was used to monitor the pattern of gene expression. Normalized data sets from microarray were surveyed in order to identify to find novel cell cycle-associated genes.

The cell cycle of HEK293T cells cannot be synchronized by serum starvation

In order to identify a suitable cell line for these studies, an experiment was performed to synchronize HEK293T by serum starvation. HEK293T cells were starved in 0.2% serum for 48 hours (h) and re-fed with 10% serum. 20 h following the refeeding with serum, cells were analysed by flow cytometry and compared to cells, which had been cultured in 10 % serum throughout the experiment. There was little evidence for the synchronization of the HEK293T population (Figure 2.3).

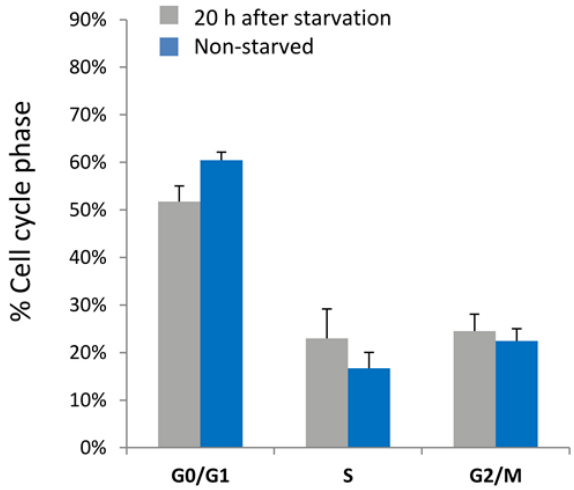


Figure 2.3: Experiment to show synchronization of HEK293T cells using serum starvation. Cells were serum starved for 48 h, re-fed with 10% serum and fixed after 20 h. Three replicates were performed for starved and non-starved control. Cells were stained with PI and analysed by flow cytometry. Percentage of cells in each phases of cell cycle is shown in graph. Mean % cell cycle phase ± standard deviation. Error bars represent standard deviation. Results based on biological repeats n = 3.

20 h after serum deprivation, the proportion of the cell population at different phases of the cell cycle i.e. G_0/G_1-, S-, or G_2/M-phase, were not statistically different compared to the non-

starved cells (Figure 2.3). This suggests that serum deprivation cannot provide a major population of HEK293T cells with a synchronized cell cycle after serum refeeding.

Synchronization of THP-1 by serum starvation

We next examined the response to serum-deprivation from THP-1 cells to serum refeeding. THP-1 cells were starved by placing them in medium with 0.2% serum for 48 h and were then re-fed in medium containing 10% serum. Flow cytometry was used to monitor the shifting of cell cycle from G_0/G_1- to S-phase at time points 0, 6, and 24 h after the serum refeeding in this experiment. Cells were cultured in the medium with 10% serum served as the control group (non-starved) (Figure 2.4).

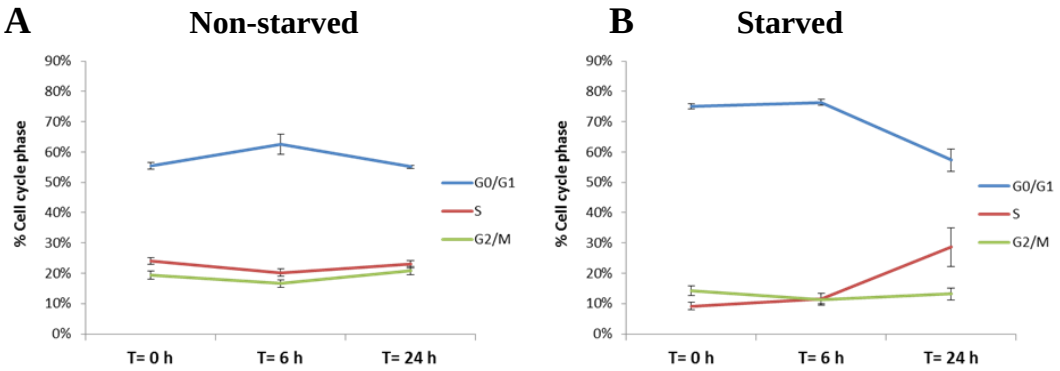


Figure 2.4: Synchronization of THP-1 cells. Percentage of control non-starved **A)** and starved **B)** cells in each phase of the cell cycle at t = 0, 6, and 24 h after the medium was renewed (mean % cell cycle phase ± standard deviation). Results based on biological repeats n = 3.

In Figure 2.4A, the percentage of non-starved cells in G_0/G_1-phase at t = 0 h showed no difference when compared to the percentage of cells at 24 h. Figure 2.4B shows that the percentage of cells in G_0/G_1-phase had decreased at t = 24 h compared to t = 0 h (T test $P < 0.01$). In addition, the percentage of cells in S-phase showed an increase in population at t = 24 h when compared to t = 0 h ($P < .05$). This suggests that starved cells responded to the re-feeding of serum, and that approximately 16% of the cells had re-entered the cell cycle by 24 h post-feeding. This indicates that the population of non-starved cells has similar cell cycle phase profiles at different time points.

Serum deprivation can provide a major population of NHDF cells with a synchronized cell cycle

As it was not possible to effectively synchronise HEK293T or THP-1 cells, an experiment was performed to synchronize NHDF cells. Figure 2.5 shows a profile of a major group of 48 h serum starved NHDF cells released from a quiescent stage following the time points at 6 h intervals after refeeding with 10 % serum. Compared with the proportion of control, starved cells responded to serum refeeding and then migrated from G_0-phase to mitotic activity. Between 18 and 24 h, a large population of cells shifted to enter the mitotic activity (Figure 2.5B).

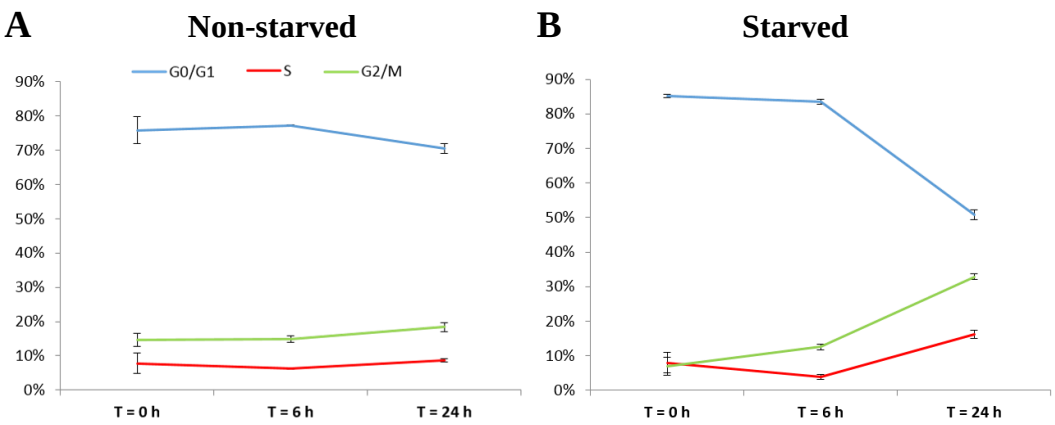


Figure 2.5: Synchronization of NHDF cells. A) Percentage of control non-starved cells in each phase of the cell cycle at t = 0, 6, and 24 h following the serum refeeding (mean % cell cycle phase ± standard deviation). The percentage of cells in G_0/G_1-, S- or, G_2/M-phase showed no significant difference at different time points. **B)** Percentage of cells at a particular phase of the cell cycle at t = 0, 6, and 24 h after the medium was replaced with 10% FCS containing medium (mean % cell cycle phase ± standard deviation). There was a 34.3 % decrease in the number of cells at G_0/G_1-phase at 24 h when compared to t = 0 h ($P < 0.01$) whilst the percentage of cells in G_2/M-phase increased by (22.3%) at t = 24 h when compared to t = 0 h ($P < 0.01$). Results based on n = 3.

Figure 2.6 demonstrates evidence of the migration of synchronous cell population responded to serum re-feeding and re-entered the cell cycle. Synchrony of cell populations responded to the serum refeeding was determined by flow cytometry at 6 h intervals over 48 h. In comparison with the percentage of non-starved cell populations, starved cells responded to re-supplementing the medium with 10% serum and then released from G_0/G_1-border across S-phase arrived G_2/M-border between 18 and 24 h.

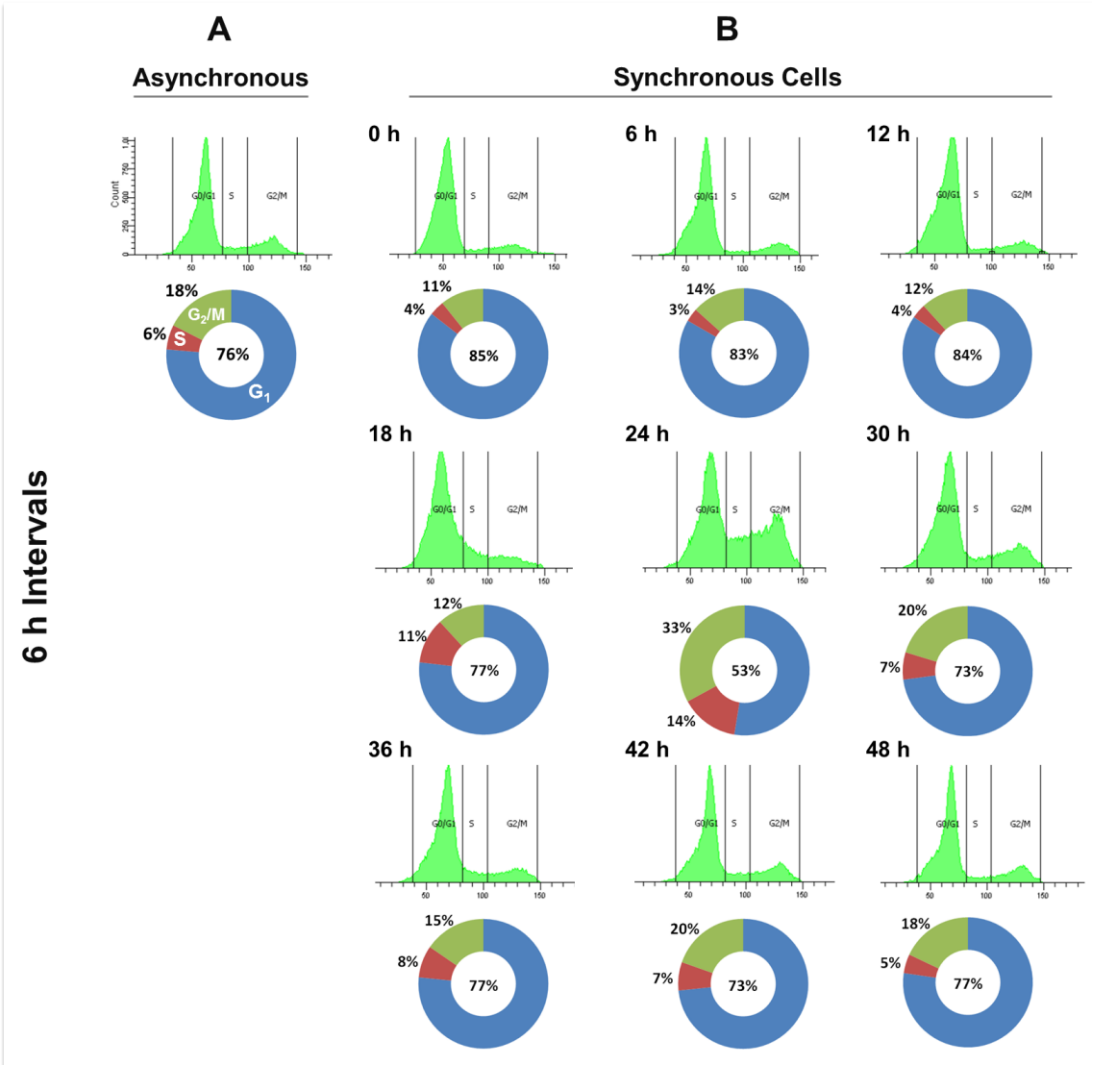
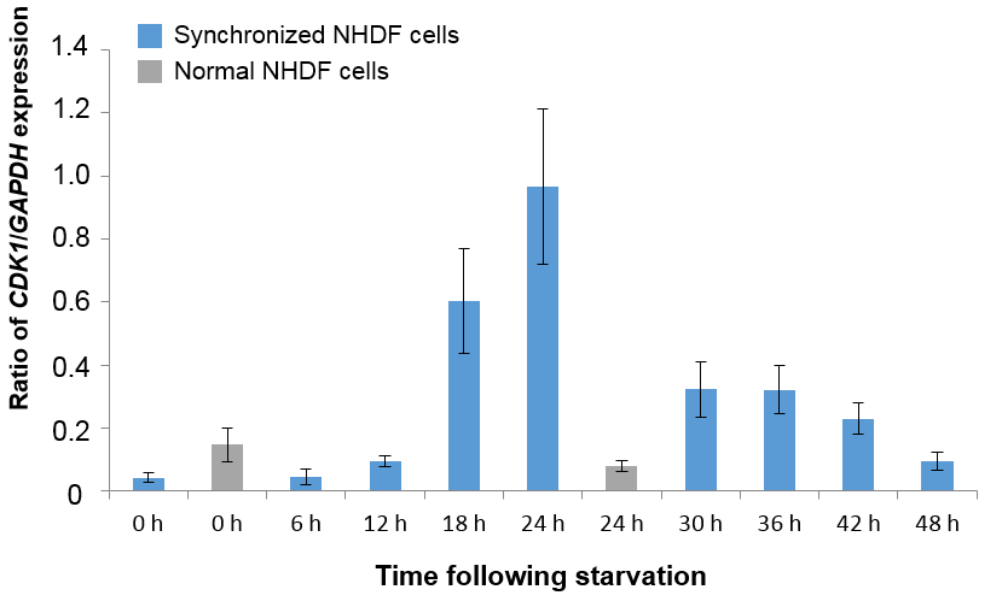


Figure 2.6: Arrest and release of NHDF cells. **A.** Flow cytometry profile of asynchronous NHDF cells cultured in medium with 10% serum. **B.** NHDF cells arrested at the G_0/G_1-phase by serum deprivation for 48 h and released by 10% serum refeeding were monitored at 6 h intervals in first cell cycle experiment and 2 h intervals in the second. Pie charts are used to visualize the % of cells in each phase of the cell cycle at different time points (Blue: G_0/G_1-phase, Red: S-phase, G_2- and M-phase is showed by green).

RT-qPCR of mRNA expression of genes up-regulated in M phase

A



B

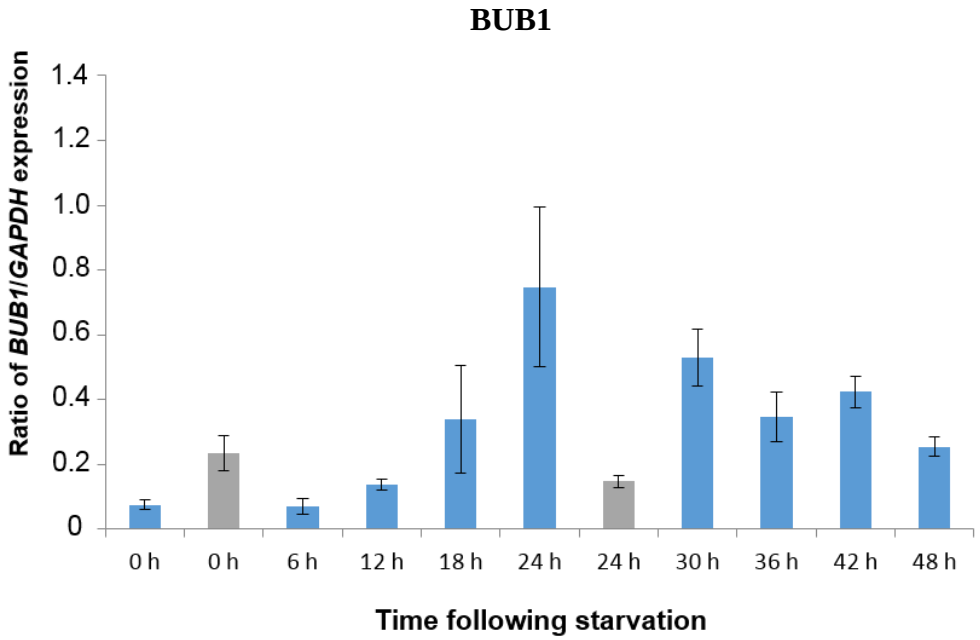


Figure 2.7: mRNA expression profiles of **A**) *CDK1* and **B**) *BUB1* following serum refeeding. Specific Primers for *CDK1* and *BUB1* were used to analyse the level of their expression in partially synchronous NHDF cells at 6 h intervals. Samples were harvested from cells following the 10% serum refeeding at 6 h intervals after 48 h 0.5% serum starvation and analysed by RT-qPCR. Controls were samples harvested from cells cultured in complete medium. Error bars represent standard deviation (n=3).

Primers specific to M-phase regulated genes *CDK1* and *BUB1* were used to detect the expression of these mRNAs by qPCR (Figure 2.7). *CDK1* is involved in the regulation of eukaryotic cell cycle, including modulating the centrosome cycle between G_2- and M-phase [44]. *CDK1* expression increases dramatically at 18 h and at 24 h following 10% serum refeeding.

BUB1 is involved in the regulation of the spindle checkpoint at M-phase [217]. *BUB1* expression peaked at 24 h after refeeding with 10% serum. Gene expression profile of *BUB1* gradually increased from 18 h, then gradually reduced from the peak expression at 24 h to the end of time course monitoring at 6 h intervals. Gene *BUB1* expression appears to increase slightly again in the latter part of the time course (T= 36 to 48 h) (Figure 2.7B). Both *CDK1* and *BUB1* gene expression profile confirm that partial synchronization has been achieved.

Normalization of microarray dataset

Before the data analysis, the quality of raw microarray data must be assessed to ensure the integrity of the data and then normalized to minimize technical variance between samples. To discover potential problems in quality with in the array samples, the chip pseudo-images visualization with probe level model of array samples at 2 h and 6 h intervals were used in this quality assessment.

Pseudo-images of microarray chips are useful for detecting special artefacts on arrays that could potentially result in quality issues. Samples with the label ‘A EXP 1 h’, and ‘A EXP 4 h’ were found to have artefacts (Figure 2.8A). In addition, arrays processes by Affymetrix GeneAtlas® System were found to be brighter than Affymetrix GeneTitan® System suggesting that there was technical variation between two systems. The difference between the two data sets can also be observed in the box plots and histogram pre-normalisation (Figures 2.8B-E).

Figure 2.9 shows a heat map generated by Pearson correlation clustering where two major clusters of array samples at 2 and 6 h intervals were placed. For the array samples collected at 6 h intervals, the cluster analysis placed them into 6 different groups i.e. group A to F. Group A includes samples harvested at 30 and 36 h. Group B includes samples harvested at 24 h after serum refeeding and control. Group C includes two array samples harvested at 18 h and one at 42 h. Group D includes two samples harvested at 48 h and one at 0 h. Group E have two samples harvested at 6 h and another two samples harvested at 12 h. Group F includes two samples harvested 6 h and another two samples harvested at 0 h.

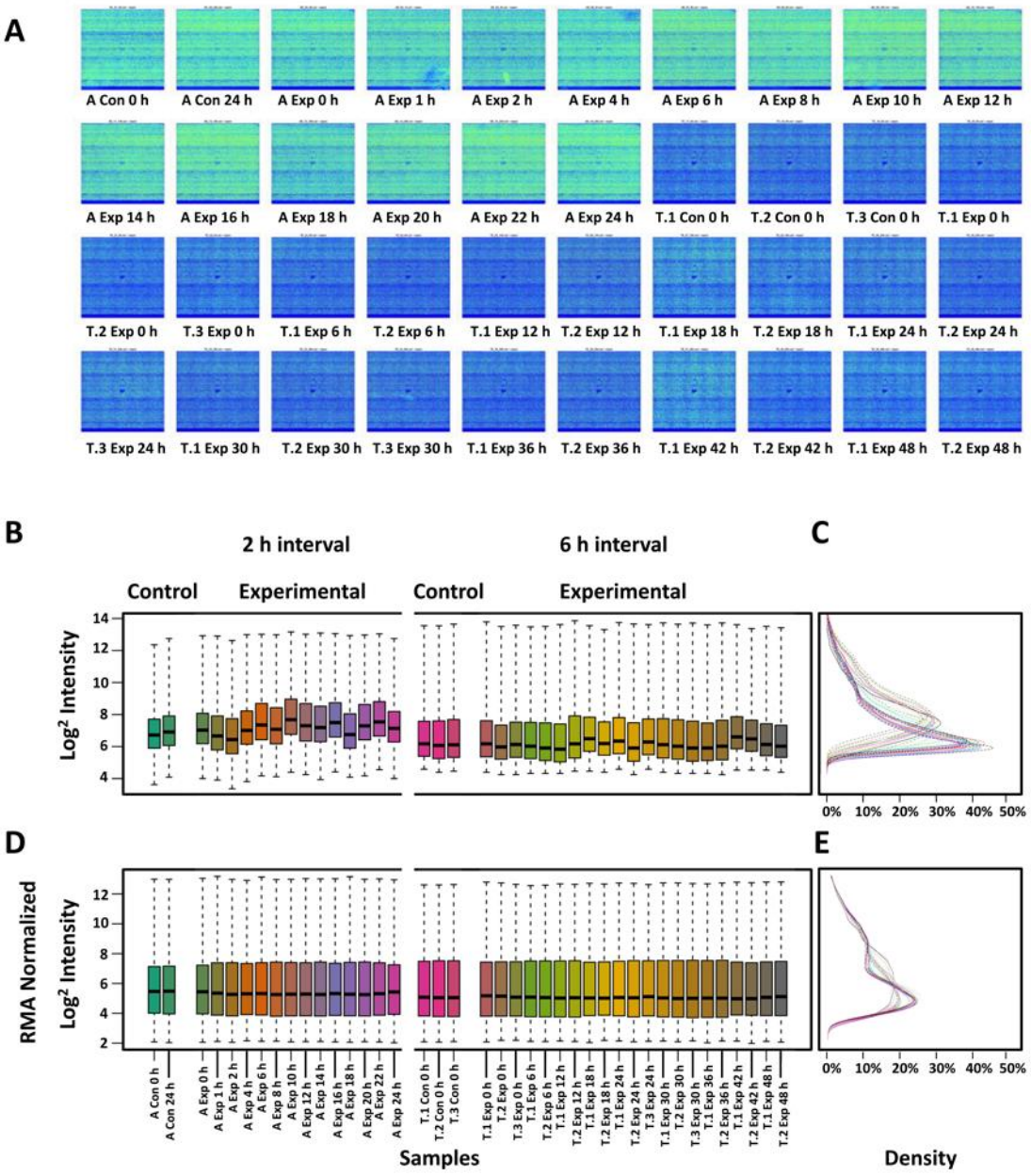


Figure 2.8: Visualization of microarray dataset and probe reading intensities. Two array experiments (samples harvested at 2 and 6 h intervals) were performed and analysed by Affymetrix® Human Gene ST 1.1 Array and then performed by Affymetrix GeneAtlas® System and Affymetrix GeneTitan® System respectively. **A)** Pseudo-image of raw data visualized as a probe level model with heat colours. **B)** Raw data of Log^2 density boxplot. **C)** Smoothed histogram of probe reading levels at log^2-intensities versus the density of probes from raw data. **D)** Log^2 density boxplot of array samples after RMA normalization. **E)** Smoothed histogram of array samples after RMA normalization.

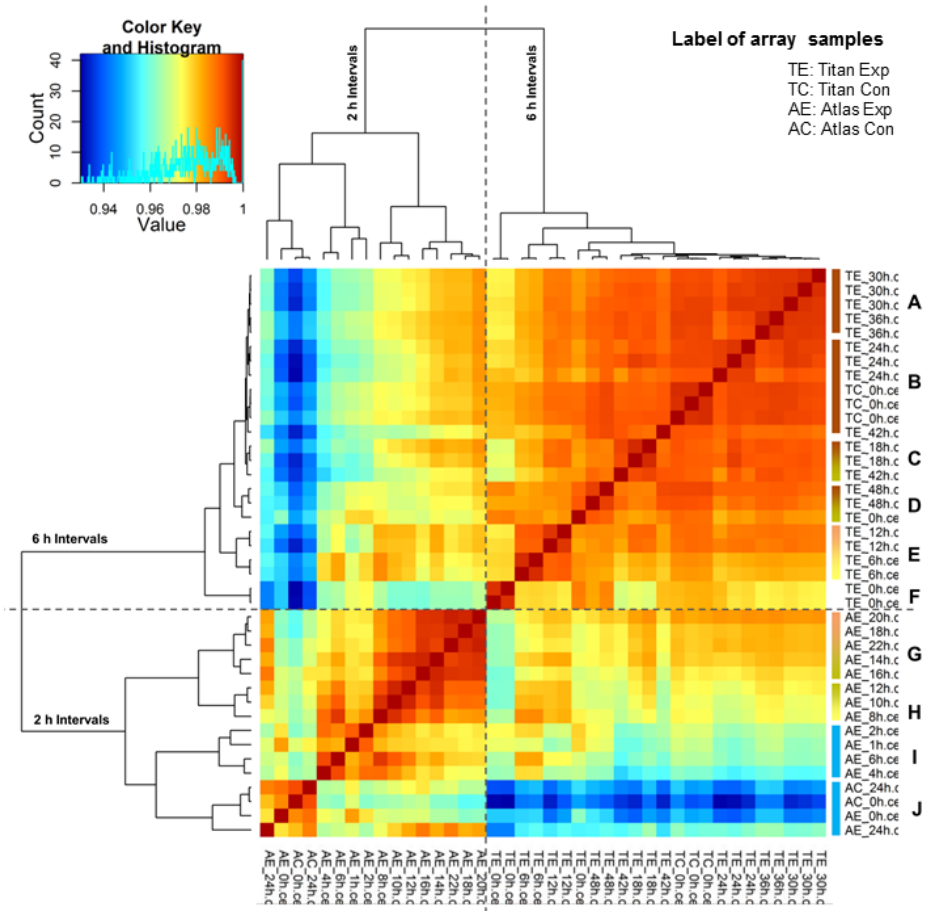


Figure 2.9: Heat map representation of the distance between RMA normalized array samples of log^2 probe intensities clustered by Pearson correlation. Samples labelled starting with AE are array samples harvested at 2 h intervals; AC means array samples harvested at 2 h intervals as the control group. Samples labelled starting with TE indicates array samples harvested at 6 h intervals as the experimental group; AC means array samples harvested at 6 h intervals as the control group.

For the second array experiments, samples were harvested at 2 h intervals (Figure 2.9). Array samples were clustered into four major groups. Group G includes array samples collected from 14 to 22 h post serum refeeding (Figure 2.9). Group H includes the array samples collected from 8 to 12 h. Group I includes samples between 1 and 6 h. Group J includes samples harvested at 0 and 24 h after serum refeeding and two controls (Figure 2.9).

PCA analyses show the grouping of samples according to array intensities before and after RMA normalisation (Figure 2.10). For array samples collected at 6 h intervals, the array intensity of samples with repeats at 0, 12, 18, 36, and 42 h were variously distributed, suggesting the diversities between samples harvested at the same time point (Figure 2.10A). RMA normalisation minimalized the diversity of array intensity of repeat sample at 0, 12, 18, 36, and 42 h (Figure 2.10B).

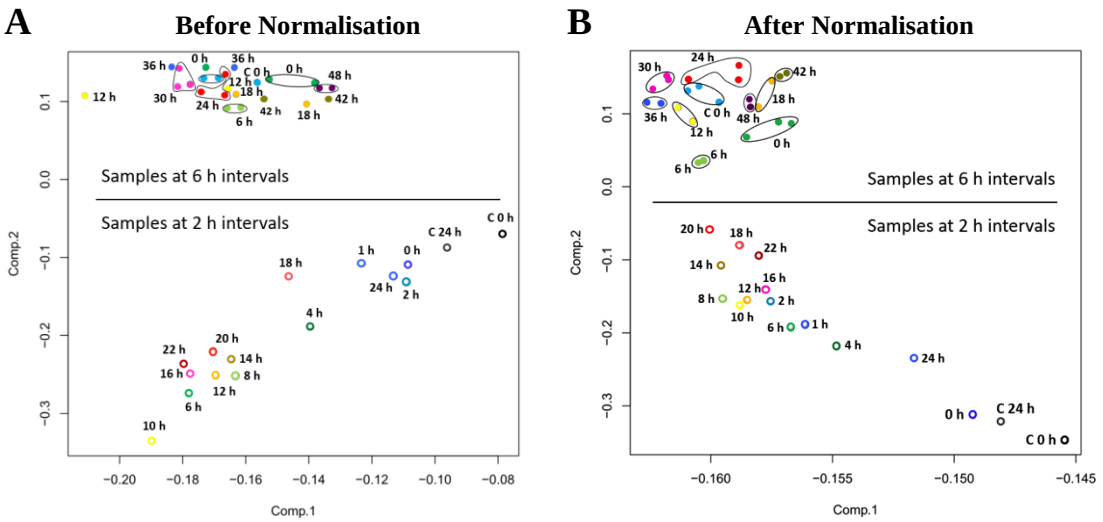


Figure 2.10: Principle component analysis (PCA) using array probe reading intensities at 2 or 6 h intervals A) before and B) after RMA normalisation. 'C 0 h' and 'C 24 h' represent the samples of control harvested at 0 and 24 h.

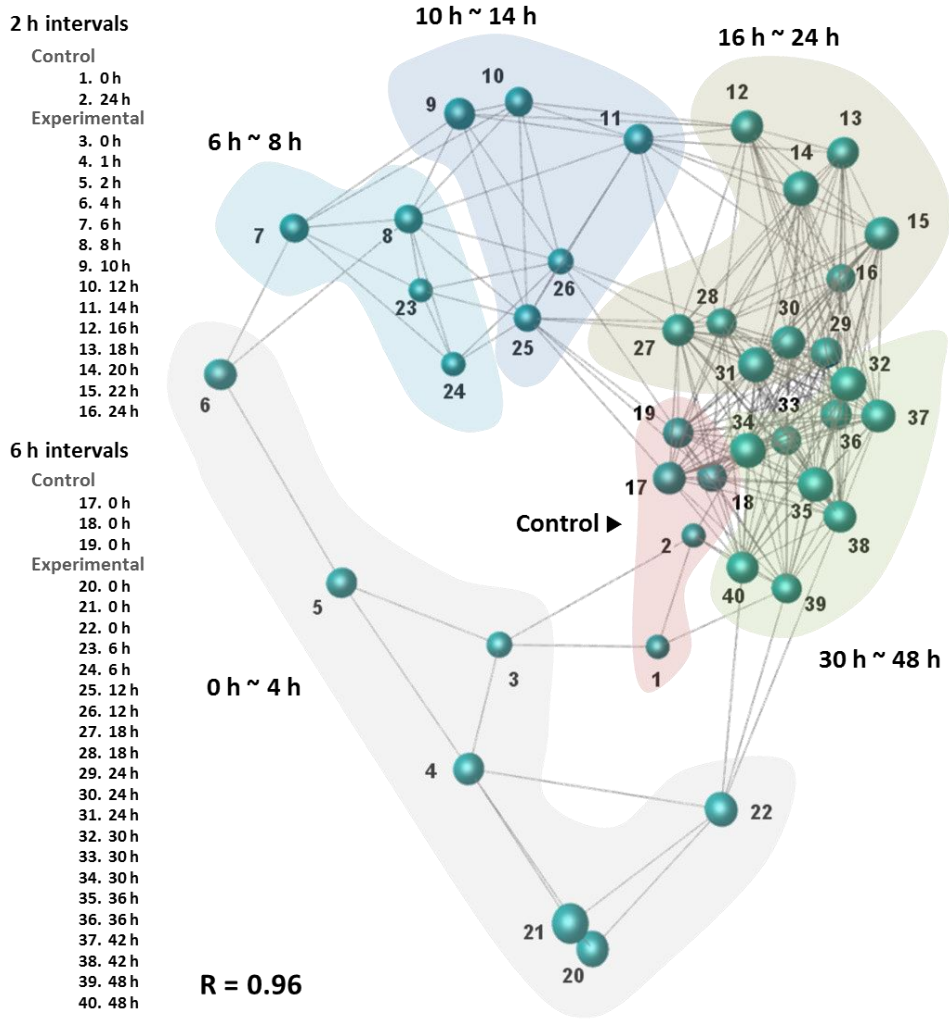


Figure 2.11 Using network analyses to determine the continuity between 40 arrays following a sequence of time point disregarding uninteresting transcripts.

For the samples harvested at 2 h intervals, PCA analysis suggests that samples were randomly distributed suggesting the probe reading intensities of each array is diverse (Figure 2.10A). In contrast, RMA normalised array intensities shows an ordered position following a sequence of time points from 0 to 24 h, which means that sample harvested at one time point was similar between its neighbour time points. This suggests that the RMA normalisation minimised the diversity between samples harvested following an ordered time points e.g. samples harvested at 8, 10, and 12 h were closed to each other (Figure 2.10B).

In Figure 2.11, network analysis was used to perform quality assessment of array samples composed of transcripts associated with the cell cycle. As can be seen, array samples from two time course experiments, across similar periods of each time point, were placed near to one another i.e. G_0: time points at early stage of the cell cycle from 0 to 4 h, G_1 early stage: t = 6 ~ 8 h, G_1 late stage: t = 10 ~ 14 h, S, G_2 and, M stage: t = 16 ~ 24 h, time points after t = 24 h, and control. The array samples from two individual time course experiments, composed of cell cycle-associated transcripts, consistently follow a sequence of time points, suggests that the pipeline of data mining is reliable and robust.

Identification of cell cycle-associated genes by network structure analysis

After quality assessment, probes with normalised expression value of < 35 (normal scale) were filtered out, resulting in a total of 21,842 transcripts out of 33,298 evaluations (Figures 2.11A and B). Using network analysis in BioLayout *Express*^{3D}, a number of batch effects were found using a MCL-inflation value of 3.0 and r = 0.85. 4,400 (20%) transcripts, correlated with the pattern of batch effect, were filtered out, resulting in 17,442 transcripts (Figures 2.12B and C). Negative and positive controls of probe reading of array were removed, resulting in 14,882 transcripts remaining in the dataset (Figures 2.12C and D). A graph containing 10,184 nodes and 270,338 edges was generated based on the correlation cut off r = 0.85. The minimum size of sub-graph components included in the network was 6. The MCL-inflation value was set at 1.7 with the smallest cluster allowed at 6 (Figure 2.12E).

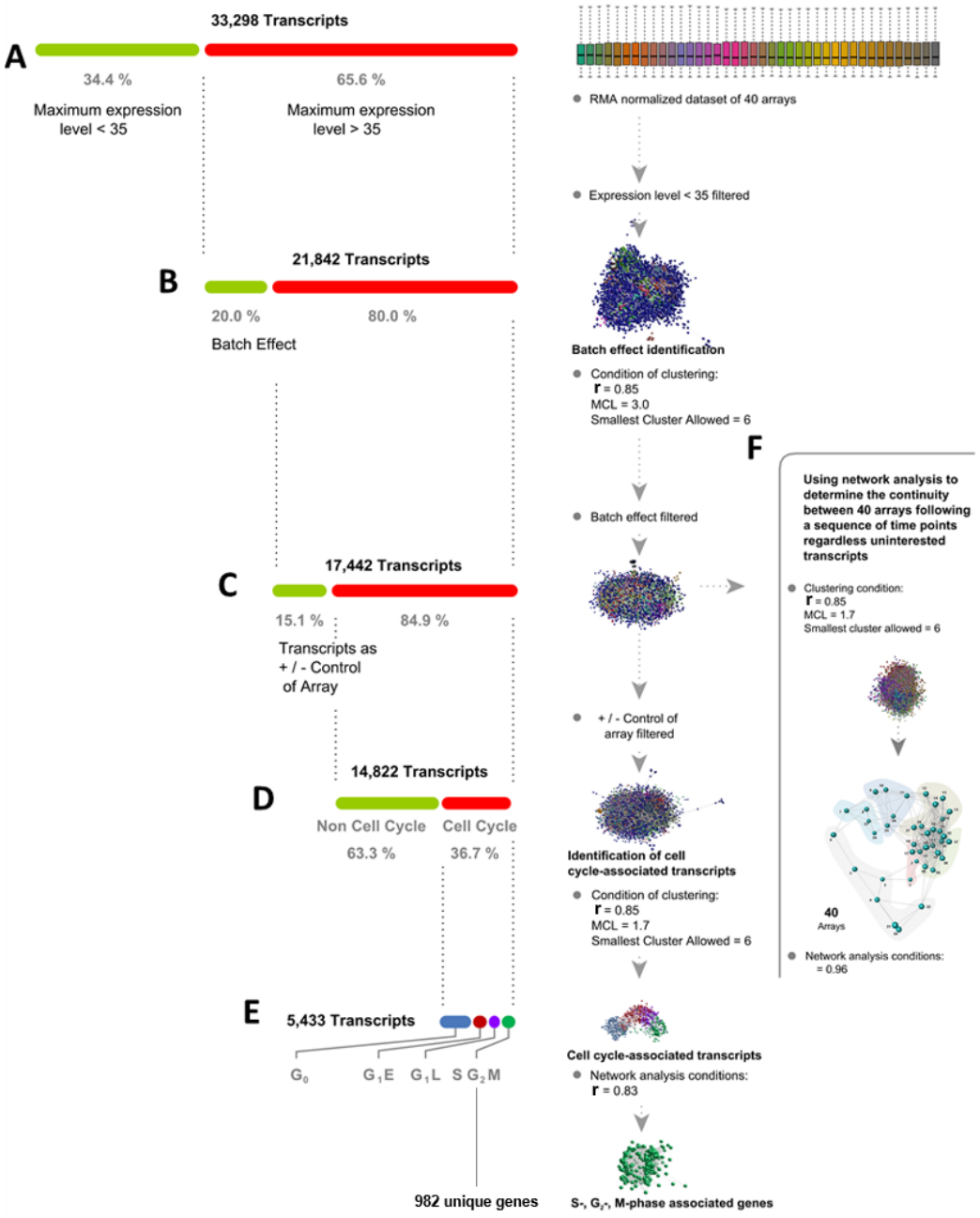


Figure 2.12: Identification of genes with putative or unknown function clustered with the cell cycle-associated genes. A) Transcripts with probe reading levels at natural scale of transcripts below 35 (natural scale) were filtered, resulting in 21,842 transcripts in the dataset. **B)** A network structure composed of clusters of gene expression profile with batch effect was determined and filtered under the condition of r = 0.85, MCL-inflation value = 3.0, and smallest cluster number = 6, resulting in 17,442 transcripts. **C)** Transcripts as negative and positive control of array were filtered, resulting in 14,822 transcripts. **D)** Cell cycle-associated clusters i.e. non cell cycle-associated transcripts (63.3 %), G_0-phase associated genes (16.7 %), G_1-phase associated genes (12.8 %), and S-, G_2/M-phase associated genes (7.2 %), were identified under the conditions of r = 0.85, MCL-inflation value = 1.7, and smallest cluster number = 6. **E)** 982 unique genes up-regulated during S-, G_2-, and M-phase were identified. **F)** Array dataset composed of the cell cycle-associated transcripts clustered from the dataset at step C was assessed by network analysis at r = 0.96 ensuring the integrity between array samples at same time point.

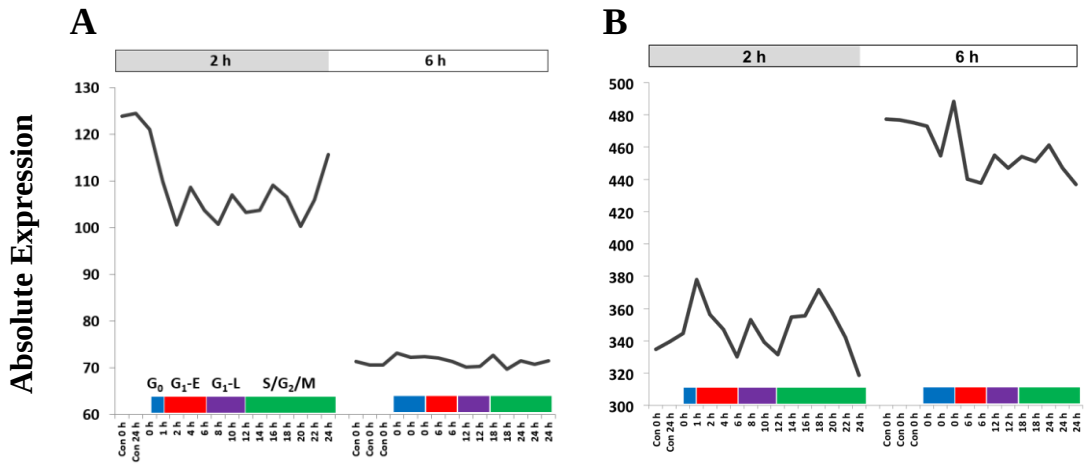


Figure 2.13: Clusters due to batch effect. Mean expression profiles of two types **A)** and **B)** of identified batch effect-associated clusters are calculated from the expression levels of all transcripts within the given clusters. Expression levels are plotted across different time-points of two experiments sampled at 2 h (plots with grey back ground) or 6 h (plots with white back ground) intervals of partially synchronized NHDF cells.

Overview of temporal expressed genes in one cell cycle

Following the removal of genes whose expression was not associated with the cell cycle, around 5,000 genes were placed in clusters and grouped based on four major waves of gene expression levels (Figure 2.12). To get an overview of the patterns of gene expression level at time points after serum refeeding, snapshots of the real-time gene expression level following a sequence of time points were animated by BioLayout *Express*^{3D} at 2 h intervals (Figure 2.14). The higher the expression level of genes, the stronger the red colour is displayed; the lower the expression level, the weaker the red colour (stronger the blue) and the smaller the node (Figure 2.14).

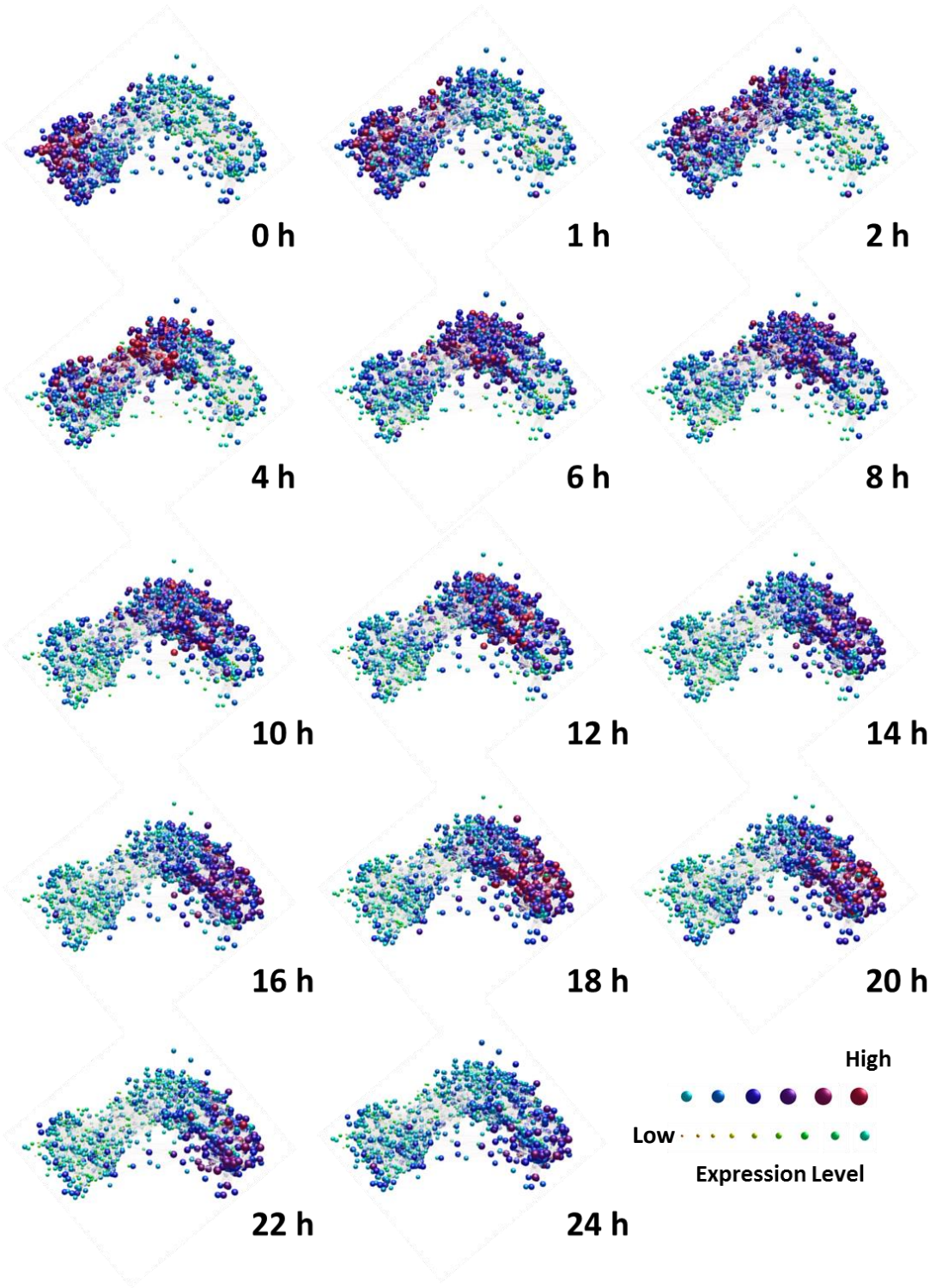


Figure 2.14: Animation of cell cycle-associated gene expression levels at 2 h intervals following the serum refeeding. Nodes represent up-regulated transcripts and are coloured according to their expression level. Edge represents correlation between nodes.

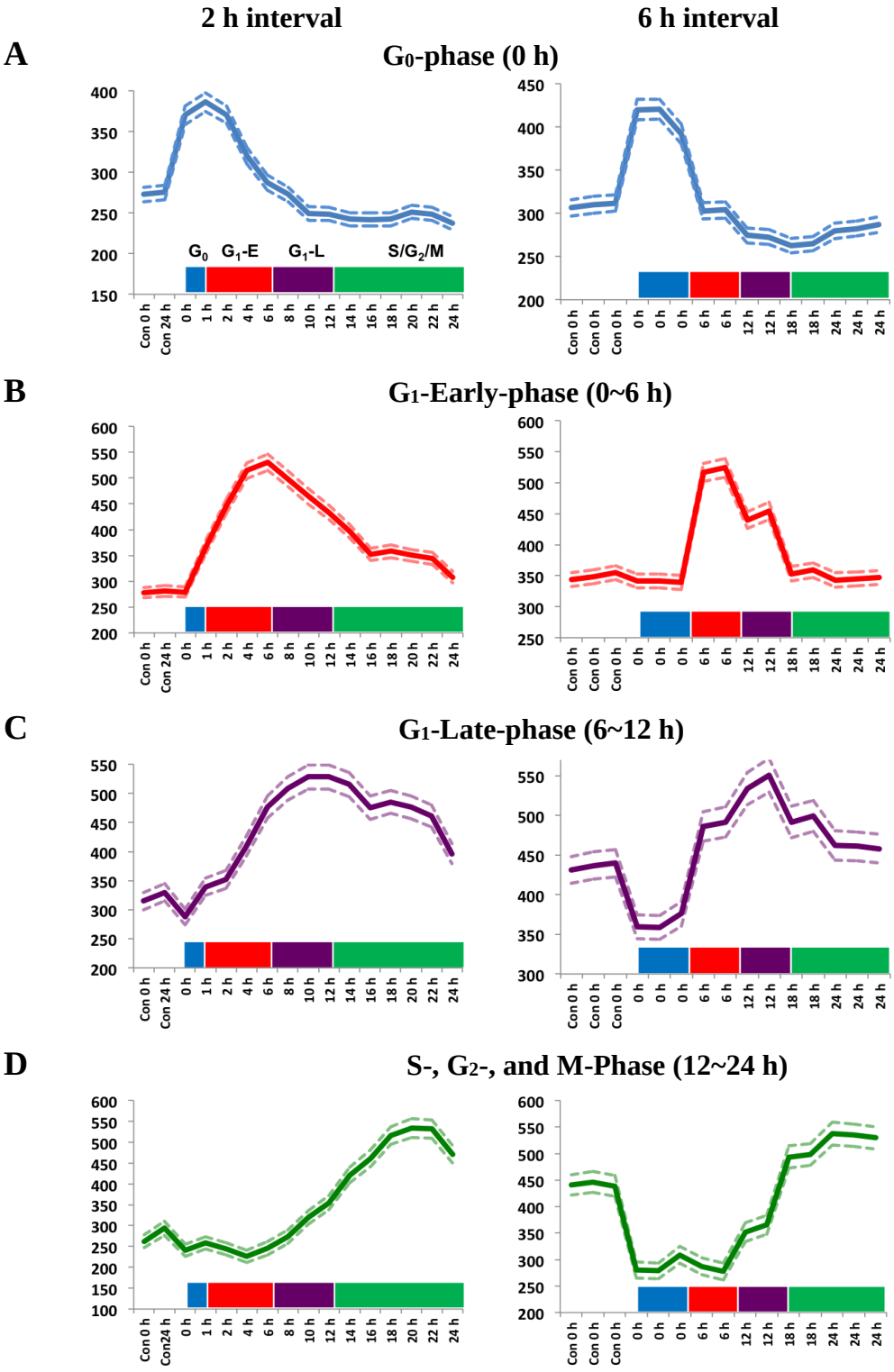


Figure 2.15: Four major waves of average clustered cell cycle-associated gene expression. Gene expression profile in two array data sets (array samples harvested at 2 and 6 h intervals) co-expressed with the cell cycle-associated genes were identified by network structure analysis classifying into **A)** G_0-phase, **B)** G_1-Early-phase, **C)** G_1-Late-phase, and **D)** S-, G_2-, and M-phase. Dashed lines represent the variation in signal.

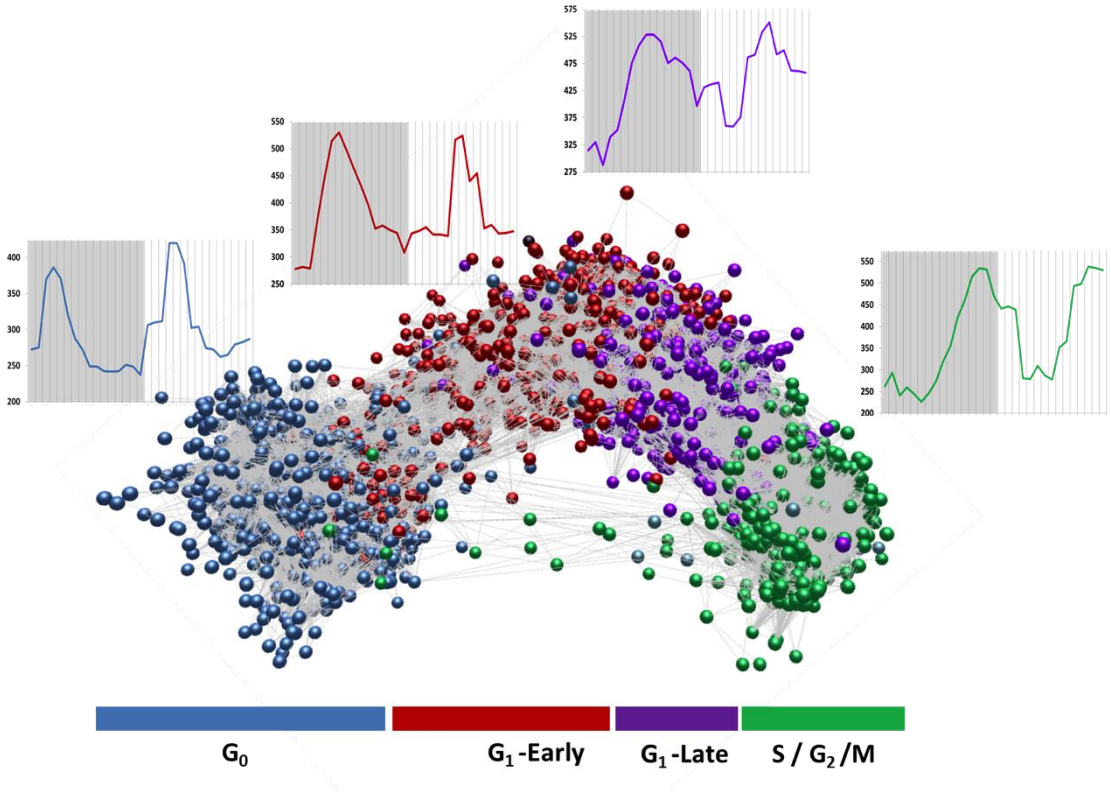


Figure 2.16: Cluster analysis of time-course data for synchronized NHDF cells following serum refeeding. Transcripts of array samples were harvested from synchronized NHDF cells following serum refeeding in two repeated time-course experiments at 2 and 6 h intervals. Gene expression was measured across a sequence of time points on Affymetrix Human Gene 1.1 ST Arrays and a network graph of the data was built by BioLayout *Express*^{3D}. The network was filtered to display only relationships with the cell cycle, resulting in a graph with 5,443 nodes (transcripts) connected by 303,403 edges. The network composed of cell cycle-associated transcripts was then clustered using the graph-based clustering algorithm MCL, set at an inflation number of 1.7. Clusters of nodes sharing similar expression feature associated with different phases of the cell cycle were given the same colours.

Based on the validation of known regulatory genes placed in clusters and their expression periodicity consistent with the cell cycle, it is suggested that construction of co-expression networks in BioLayout *Express*^{3D} has identified 4,528 unique genes with an expression periodicity that correlates with cell cycle of the cells including 1780 G_0-phase clustered genes, 885 genes up-regulated in the early stage of G_1-phase, 881 genes expression peaked at the late stage of G_1-phase, and 982 genes are associated with the genes in S-, G_2-, and M-phase (Figure 2.16).

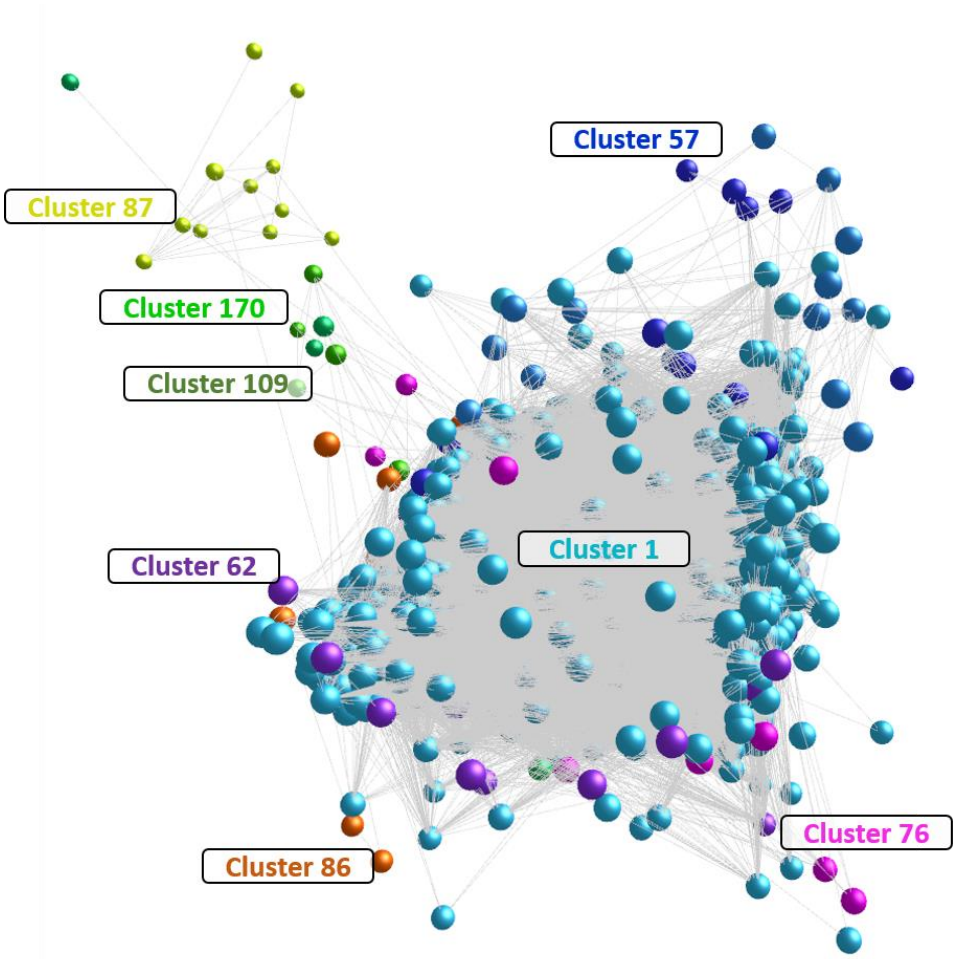


Figure 2.17: 8 sub-clusters of the major cluster associated with S-, G_2-, and M-phase. Number of clusters were categorised according to the size of sub-clusters belonging to the entire network structure. The size of each sub-cluster determines the order of their numbers. For example, the number of nodes in cluster 01 is bigger than cluster 170. For more details of sub-clusters please see table 2.2.

To further investigate genes that clustered with known cell cycle-regulated genes, figure 2.17 was used to demonstrate the network structure of a major cluster composed of 8 sub-clusters in different colours. Cluster 1 (901 unique genes) included the major population of genes' expression profile associated with the S-, G_2-, and M-phase. Cluster 57, 60, 62, 76, 86, 87, 109, 170, have 8, 15, 12, 11, 10, 11, 9, 7 unique genes respectively.

Functional annotation of the cell cycle-associated genes by GO slim mapping

In Figure 2.18, the number of S/G_2/M-phase genes categorised to the terms of the cell cycle and cell proliferation is higher than the number of genes associated with other cell cycle phases. In contrast, the number of G_0-phase genes is higher than the number of genes associated with other phases in the cell cycle i.e. signal transduction, multicellular organismal development, metabolic process, ion transport, lipid metabolic process, cell differentiation, carbohydrate metabolic process, transport, and protein modification process.

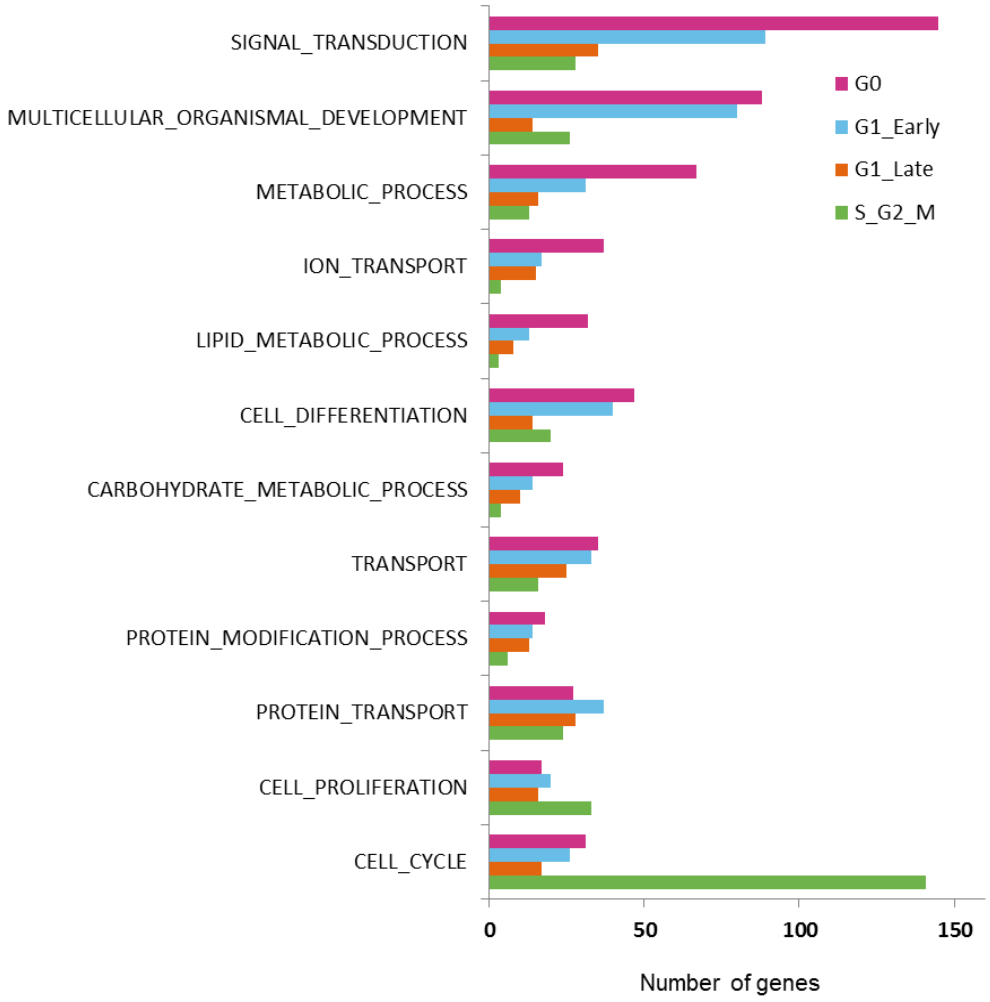


Figure 2.18: Profiling of the list of cell cycle-associated genes by GO slim mapping. A cut down-version of GO ontologies was used to give a broad overview of functional categorized list of cell cycle-associated genes. Clusters of genes were categorized into different terms of biological processes.

For example, the number of categorized genes in S- and, G_2/M-phase is only one fifth of the number of genes up-regulated in G_0-phase. The signal transduction in GO slims includes the

cellular process where a signal is transferred to initiate the change of the state of cells, such as the regulation of transcription or metabolic process. This suggests that at the level of biological process, starved NHDF cells respond to the serum refeeding and re-enter a synchronized cell cycle.

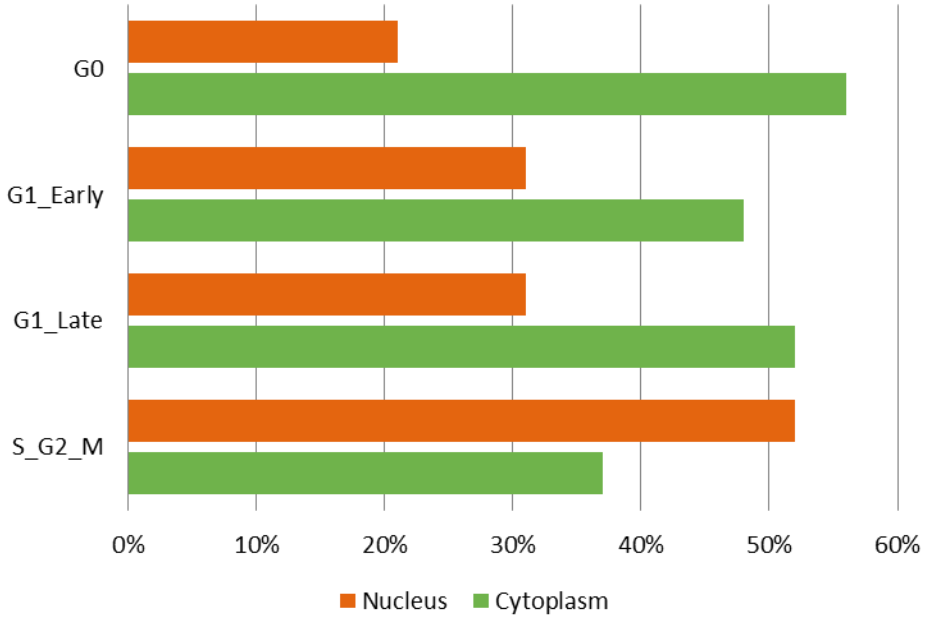


Figure 2.19: GO slims mapping of the cellular components of cell cycle-associated gene. The list of cell cycle-associated genes was categorized by a cut-version of GO ontologies based on cellular components. Terms of localised gene products in different phases of the cell cycle were simply classified into cytoplasm and nucleus.

Using GO slims, the list of genes up-regulated in the cell cycle were mapped providing a profile of cellular location of gene products. The distribution of nucleus localised gene products in G_0-phase is only around one third of gene products in cytoplasm (Figure 2.19). The distribution of gene products in G_1-Early and late phase comprises nearly two thirds of the gene products located in cytoplasm. In contrast, during S-, G_2-, and M-phase, protein expressed by genes in nucleus is around one and half times greater than in the cytoplasm.

GO term enrichment analysis of clustered genes

To further investigate clustered genes as their expression periodicities are in association with the cell cycle, they were investigated by GO term enrichment analysis i.e. genes that are up-regulated in G_0, G_1-early, G_1-late, and S/G_2/M-phase. GO term ID enriched for associated genes, and their p-values were visualised in Figure 2.20. Details of enriched terms and genes were summarised in Table 2.2.

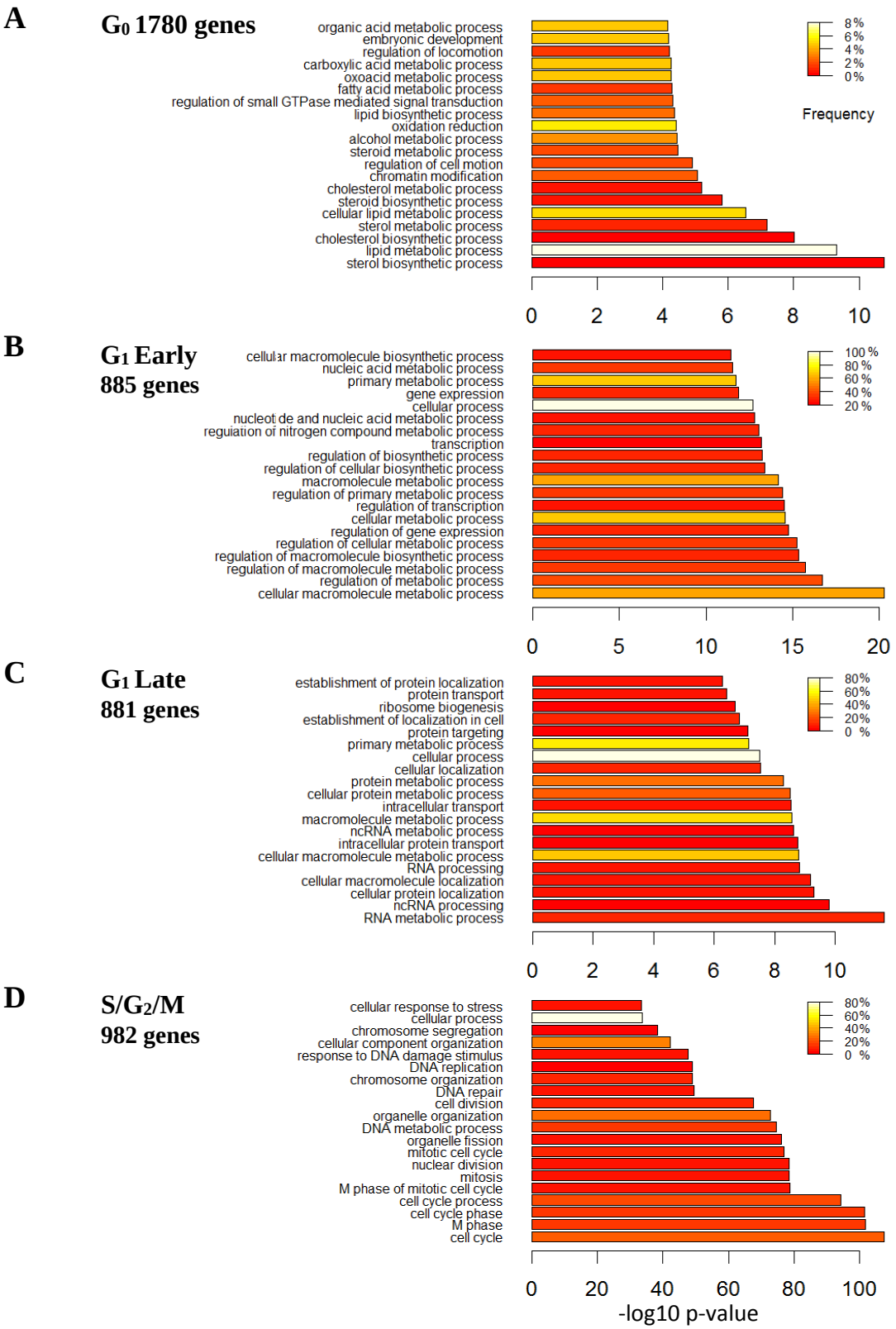


Figure 2.20: Bar chart showing the score of the top 20 GO term enriched by the analysis of genes clustered in A) G_0, B) G_1 Early stage, C) G_1 Late stage, and D) S/G_2/M-phase groups. Colour key of frequency shows the ratio of genes, which are annotated to specific GO term against input number of genes. Horizontal bar represents transformed p-value at –log10 scale.

A

G_0 (2080 transcripts, 1780 unique genes)

Example regulatory genes: *TFDP2, CDKN1A*

Cluster: 3, 5, 8, 13, 14, 23, 24, 26, 32, 36, 41, 42, 46, 55, 61, 73, 80, 81, 100, 104, 118,125, 131, 138, 156, 157, 161, 162

GO terms: *sterol metabolic process, cholesterol metabolic process, lipid biosynthetic process*

B

G_1-E (1364 transcripts, 885 unique genes)

Example regulatory genes: *CCND1, CDK6, E2F4, E2F5*

Cluster: 2, 11, 18, 19, 20, 25, 29, 30, 31, 47, 48, 51, 56, 58, 68, 84, 98, 108, 120, 122,129, 133, 143, 149, 153, 164, 175

GO terms: *transcription, DNA-dependent regulation of transcription, regulation of RNA metabolic process*

C

G_1-L (926 transcripts, 881 unique genes)

Example regulatory genes: *CCNE1, CCNE2, CDC25A, MNAT1, ORC5, MCM, CDC6, PPP2R1B*

Cluster 7, 9, 10, 37, 45, 49, 59, 70, 74, 99, 106, 119,160, 166,167, 178,179

GO terms: *protein transport, intracellular protein transport, protein localisation in organelle, ncRNA metabolic process, ribosome biogenesis, ribonucleoprotein complex biogenesis, tRNA metabolic process, rRNA metabolic process*

D

S/G_2/M (1067 transcripts, 982 unique genes)

Example regulatory genes: *TOP2A, PCNA, MCM2, CCNA2, CCNB1, CCNB2, CDC25B, CDK1, BRIC5, BUB1, MAD2L1, CDC20, CENPF*

Cluster: 1, 57, 60, 62, 76, 86, 87, 109, 170

GO terms: *cell cycle, cell cycle process, mitotic cell cycle, cell division DNA replication initiation, G_2 DNA damage checkpoint, DNA repair, chromatin assembly, DNA recombination, spindle organization, mitotic spindle elongation, attachment of spindle microtubules to kinetochore, mitotic cell cycle checkpoint, mitotic sister chromatid segregation*

Table 2.2: Number of transcripts, number of unique genes placed in four major clusters, sub-clusters, and results of the enrichment analyses.

In Table 2.2A, GO associations of clustered genes associated to G_0-phase were enriched for the ‘sterol metabolic process’, ‘cholesterol metabolic process’, ‘lipid biosynthetic process’, and ‘isoprenoid metabolic process’, such as transmembrane 7 superfamily member 2 (*TM7SF2*) were involved in the regulation of conversion of lanosterol to cholesterol [218].

In Table 2.2B, for clustered genes up-regulated between 0 and 6 h were enriched for the terms ‘transcription, ‘regulation of transcription’, ‘DNA-dependent’, ‘regulation of RNA

metabolic process'. Major biological events are relevant to transcription activities. For example, histone cell cycle regulator (*HIRA*) is involved in the regulation of histone gene transcription [219]. Signal transducer and activator of transcription 5A (*STAT5A*) responds to cytokines stimulation (i.e. interleukin 2 (IL-2)), shifting into the nucleus as a transcription activators [220].

In Table 2.2C, genes with peak expression between 6 and 12 h were enriched for the major biological activity as post-transcriptional regulation, including the GO terms 'protein transport', 'intracellular protein transport', 'intracellular transport' and 'protein localisation in organelle'. For example, RNA binding motif protein 22 (*RBM22*) is involved in the positive regulation of RNA splicing [221]. Latent transforming growth factor beta binding protein 2 (*LTBP2*) belongs to the latent transforming growth factor (LTBP) family, which is involved in protein secretion [222]. Derlin 2 (*DERL2*) is involved in the endoplasmic reticulum (ER)-associated ubiquitin-dependent protein catabolic process [223]. This suggests that from 6 to 12 h after serum stimulation, the cell cycle regulation system progresses from the stage of the transcriptional level to the transcriptome level.

In Table 2.2D, the list of genes up-regulated between 12 and 24 h was enriched by GO analysis for terms associated with the cell cycle from the beginning of S–phase to the end of mitotic activity, such as 'DNA replication initiation', 'G_2 DNA damage checkpoint', 'DNA repair', 'chromatin assembly', 'DNA recombination', 'spindle organization', 'mitotic spindle elongation', 'attachment of spindle microtubules to kinetochore', mitotic cell cycle checkpoint, mitotic sister chromatid segregation'. Example genes of S-phase include DBF4 Zinc Finger (*DBF4*), essentially to the initiation of DNA replication with Cell Division Cycle 7 (*CDC7*) [224]. Claspin (*CLSPN*) is involved in the G_2/M checkpoint of DNA damage [225]. RAD51 paralog C (*RAD51C*) is important in activating the checkpoint kinase arresting the cell cycle for DNA damage [226]. Following S-phase, CCCTC-binding Factor (*CTCF*), a zinc finger protein, physically connects cohesion and chromatin [227]. In M-phase, Centrosomal Protein 72kDa (*CEP72*) is involved in the recruiting of key centrosomal proteins to the centrosome, providing a focused bipolar spindle which gives proper tension between sister chromatids [228]. Protein regulator of cytokinesis 1 (*PRC1*) is involved in mitotic spindle elongation and is required for cytokinesis [229]. GO term enrichment analysis suggested that genes which were placed in these clusters and expressed between 12 and 24 h following serum refeeding, were associated with the mitotic activity of the cell cycle. Evidence provided a highly reliable list of genes of interest across S-, G_2-, and M-

phase for the further study of putative novel genes e.g. the consistence of gene expression periodicity during S-, G_2-, and M-phase, identification of mitotic-associated regulatory genes, and GO term enrichment.

Identification of genes up-regulated in the cell cycle

The four major waves of gene expression were observed and well correlated with the different phases of cell cycle (Figures 2.15 and 2.16). Using known cell cycle-regulation genes provided by the pathway browser 'REACTOME' and previously published literature, important cell cycle-associated genes were identified in four main clusters of their expression periodicity correlated with the cell cycle.

Genes up-regulated in G_0-phase

Using the information provided by Hilary A. Coller *et al*, 2006, 29 known G_0-phase associated genes in this main cluster were identified, suggesting that these genes were associated with a quiescent stage [22] i.e. important known G_0-phase regulatory genes cyclin-dependent kinase inhibitor 1A (*CDKN1A*) [230] and cyclin-dependent kinase inhibitor 2B (*CDKN2B*) [231] were both the inhibitors of G_1-phase CDKs.

Genes up-regulated during the early stage of G_1-phase

Following G_0-phase, two waves of gene expression correlating with early and late stage in G_1-phase were identified. Genes placed in G_1-Early stage-associated clusters showed an increased expression level, peaking between 0 and 6 h (Figure 2.15B). In these clusters of the early stage of G_1-phase (G_1-ealry-phase), important cell cycle-regulated genes were identified as known G_1-associated genes expressed following the serum refeeding i.e. regulatory genes cyclin D1 (*CCND1*) [232] and *CDK6* [233] both play key positions during the G_1-phase cell cycle control system. E2F transcription factor 4 (*E2F4*) and E2F transcription factor 5 (*E2F5*) are both involved in the positive regulation of transcription in the early stage of the cell cycle [234].

Genes up-regulated in the late stage of G_1-phase

Clusters of genes with peak expression levels between 6 and 12 h were classified as G_1-late stage associated genes (Figure 2.15C). In these clusters, important cell cycle-regulated genes were identified, determining that these clusters were associated with G_1-phase, such as cyclin E1 (*CCNE1*) [232] and cyclin E2 (*CCNE2*) [235]. Example regulatory genes in the late stage of G_1-phase, such as CDC25A which activates CDK2 in the progression from G_1- to the early stage of S-phase [236]; MNAT CDK-activating kinase assembly factor 1 (*MNAT1*),

involved in the stabilization of cyclin H/cyclin dependent kinase 7 (CDK7) complex, providing a functional CDK-activating kinase enzymatic complex [237]; Origin recognition complex, subunit 5 (*ORC5*) involved in the early stage of DNA replication, recruiting essential factors, such as CDC6 [238] and MCM [239]; Protein phosphatase 2, regulatory subunit A, beta (*PPP2R1B*), involved in the regulation of extrinsic apoptotic signalling pathway [240].

Genes up-regulated in S-, G_2-, and M-phase

Clusters of genes up-regulated between 12 and 24 h following serum refeeding were identified as being associated with S-, G_2-, and M-phase (Figure 2.15D). Comparison to the cell cycle-associated gene list from the REACTOME pathway database, there were 56 S-phase regulated genes clustered with other genes. Regulatory genes, such as topoisomerase (DNA) II α (*TOP2A*), and proliferating cell nuclear antigen (*PCNA*) were used to determine the clusters. TOP2A is involved in the alternation of the topological state of DNA during transcription and replication [241].

41 cell cycle-associated genes involved in the transition between G_2- and M-phase were identified in S-, G_2-, and M-phase associated clusters; essential regulatory genes such as cyclin A2 (*CCNA2*), cyclin B1 (*CCNB1*), cyclin B2 (*CCNB2*), and cell division cycle 25B (*CDC25B*) provided the best evidence for the identification. *CCNA2* is essential in driving the cell cycle during G_2- and M-phase and involved in the regulation of the mitotic G_2-phase DNA damage checkpoint [242]. Proteins expressed by *CCNB1* and *CCNB2* both function as activators of CDK1 in mitosis [243]. CDC25B activates CDK1 in cytoplasm during S- and G_2-phase and then conducts the cells into mitosis [244].

95 genes involved in M-phase were identified, such as *CDK1*, budding uninhibited by benzimidazoles 1 (*BUB1*) and cell division cycle 20 (*CDC20*) were used to identify clusters associated with mitotic activity in the cell cycle. BUB1 is a mitotic checkpoint kinase [217]. *CDC20* revises the activation of mitotic anaphase-promoting complex activity [245]. For the rest of the example genes, *BUB1* is involved in the regulation of mitotic spindle assembly [217]; MAD2 mitotic arrest deficient-like 1 (*MAD2L1*) is involved in positive regulation of the spindle assembly checkpoint [246]; centromere protein F (*CENPF*) is involved in kinetochore assembly activity [247].

Categorization of cell cycle genes in S-, G_2-, and M-phase

Following identification of the cell cycle genes up-regulated in different phases, a functional list was developed to profile cell cycle-associated genes. In order to further survey the function of identified genes in S-, G_2-, and M-phase, defined annotations were used to classify the list of genes based on their characterization in the cell cycle. Associated terms of classification include 'unknown function' (Unknown function of genes expressed in the cell cycle), 'unrelated function' (Unrelated function of genes up-regulated in the cell cycle), 'putative association' (Putative associated function of genes involved in the cell cycle regulation), and 'known function' (Known function of genes in the cell cycle).

The key cell cycle-regulated genes were annotated as 'known cell cycle associated function'; a good example is *CCNB1*. Genes potentially involved in the biological processes in S-, G_2-, and M-phase were annotated as 'putative associated function in the cell cycle', for example ATPase family, AAA domain containing 2 (*ATAD2*) which may have a role in the modification and co-regulation of occupancy of chromatin [248]. Known genes up-regulated in S-, G_2-, and M-phase, without direct evidence of involvement, were annotated as 'unrelated function in the cell cycle', for example achalasia, adrenocortical insufficiency, alacrimia (*AAAS*) which may play a role in the development of the nervous system [249]. Genes without supportive information to indicate their roles either in cell proliferation or biological activities were annotated as 'unknown function', for example chromosome 17 open reading frame 53 (*C17orf53*).

Manual curation and functional annotation of genes

To further refine the list of S-, G_2-, and M-phase genes, genes that are highly expressed in human cells and tissues with high levels of cell proliferation were selected. Of 982 clustered genes that were up-regulated during S-, G_2-, and M-phase genes, 623 of them were found to also be highly expressed across different human cells and tissues. In addition, 83 genes that appear to be up-regulated during S-, G_2-, and M-phase were selected from our array dataset whilst they were also highly activated between various human cells and tissues. This resulted in a total number of 706 genes of interest.

Manual curation showed that 484 of these are known cell cycle-associated genes, 75 have known roles in other biological processes, 78 are genes with putative association to the cell cycle, and 69 have not previously been associated with the cell cycle or any other biological

process and potentially represent new components of the cell cycle machinery. This table was developed leading by Prof. Freeman's group (Table 2.3).

In table 2.3, the functional gene annotation categorised genes of interest into four major groups, including 1) the genes with known cell cycle function (i.e. 484 genes), 2) genes with a putative role in the cell cycle (i.e. 78 genes), 3) genes co-expressed with known cell cycle-associated genes but with poor information determining their involvement (i.e. 69 genes), and other known genes involved in non-cell cycle events (i.e. 75 genes).

Known cell cycle-associated genes

Of the genes classified as known S-, G_2-, and M-phase associated (484 genes), 113 genes are involved in the cytokinesis pathway, 99 genes are involved in chromatin organisation, 91 genes are involved in DNA repair, 58 genes are involved in DNA synthesis. In addition, 52 genes are involved in the cell cycle regulation, 36 genes are involved in centrosome maturation and duplication, 15 genes are involved in nucleotide biosynthesis, 15 genes are involved in the transcription regulation network, and 5 genes are involved in G_1/S phase transition. The following paragraphs are a summary of the population of genes involved in different biological process driving the cell cycle.

Of the 113 genes that are involved the cytokinesis pathway, one of the biggest sub-groups is the kinetochore complex pathway e.g. spindle and kinetochore associated complex subunit 1 (*SKA1*) [250]. Another sub-group is the family of kinesin-associated genes e.g. Kinesin Family Member 15 (*KIF15*) which is involved in the mitotic spindle assembly [251]. It is interesting that for the 99 genes which are involved in chromatin organisation, histone family genes occupy the biggest proportion of sub-groups including 56 genes e.g. histone cluster 1, H1a (HIST1H1A) [252].

For the known genes that are involved in DNA repair pathways, it is interesting that the biggest population of gene family is fanconi anemia (FA) associated genes. FA is a genetic disease induced by a genetic defect in a cluster of proteins involved in DNA repair. For example, fanconi anemia, complementation group A (*FANCA*) encodes a DNA repair protein [253]. Another interesting group of genes are involved in X-ray repair e.g. X-ray repair complementing defective repair in chinese hamster cells 1 (*XRCC1*) involved in the correction of defective DNA strand-break repair [254].

In addition, of those genes that are involved in DNA synthesis, there are 9 genes belonging to the minichromosome maintenance complex family e.g. minichromosome maintenance

complex component 6 (MCM6) that initiates the DNA duplication [255]. Others such as genes encoded for polymerase, replication factors, and origin recognition complexes, provide consolidated evidence to determine the rotation of cell cycle e.g. polymerase (DNA directed), α1 (*POLA1*) is involved in the initiation of DNA replication [256].

For genes involved in the regulatory network of cell cycle, the most import evidence is the up-regulation of cyclins (i.e. cyclin A2, B1, B2, and F), CDKs (i.e. CDK1, 2, and 4), determining the rotation of the cell cycle. For genes that are involved in the centrosome life cycle (i.e. 36 genes), this group includes the HAUS augmin-like complex and centrosome protein family e.g. HAUS augmin-like complex, subunit 1 (*HAUS1*) and centrosomal protein 135kDa (*CEP135*). Proteins encoded by *HAUS* family involves in the regulation of mitotic spindle assembly [257]. CEP135 is the core elements of the cartwheel structure of centrosomes [258].

Putative cell cycle-associated genes

For putative cell cycle-associated genes, 24 genes were thought to be involved in nuclear transport, 11 genes were potentially involved in apoptosis, 8 genes were potentially involved in RNA processing, 4 genes may play roles in transcription regulation, and 3 genes had putative function in ubiquitination.

Of those proteins encoded by genes, which were thought to involve in the apoptosis (i.e. 11 genes), 6 of them were located at the nucleus, the others were with poor information. For genes, which are labelled as nuclear transport, 12 of these are localised at the nuclear membrane, 5 of them are localised at the nucleus and the others are remained unknown. Other genes localised at the nucleus are associated with the biological processes of RNA processing, transcription regulation, and ubiquitination.

Co-expressed genes with unknown function in the cell cycle

Of unknown function genes that co-expressed with the known genes in the cell cycle, 56 of 69 genes were found with poor information to locate their role in the cell cycle-associated events. For the rest, 12 of them were zinc finger proteins, which are in association with the transcription but with poor information to determine their role in the cell cycle pathways. One of them is associated with the metabolism at mitochondria but has no information relevant to the cell cycle (i.e. farnesyltransferase, CAAX box, beta (FNTB)) [259].

For the genes with known function but which are not involved in the cell cycle, they were classified as unrelated function. 40 of the total 75 unrelated function genes were labelled as

having “other functions”. 20 genes were found to be involved in RNA processing. The others were found to be involved in metabolism, transcription, and ubiquitination.

Known S/G_2/M-phase protein	**484**
Chromatin organisation	99
Cytokinesis	113
DNA repair	91
DNA synthesis	58
Centrosome maturation and duplication	36
Regulation of cell cycle	52
Transcription regulation	15
Nucleotide biosynthesis	15
G_1-S phase transition	5
Putative association	**78**
Nuclear transport	24
Other	19
Apoptosis	11
Unknown	9
RNA processing	8
Transcription regulation	4
Ubiquitination	3
Unknown function	**69**
Metabolism	1
Transcription regulation	10
Transcriptional regulation	2
Unknown	56
Unrelated function	**75**
Metabolism	8
Other	40
RNA processing	20
Transcription regulation	6
Ubiquitination	1

Table 2.3: Summary of the functional annotation of 706 genes by Prof. Freeman’s group.

Serum deprivation dependent technique can provide samples with up-regulated gene expression correlated with the progression of the cell cycle

As can be seen from Figure 2.21, there are inconsistent expression features of genes in fibroblasts synchronized by different techniques, such as serum deprivation and serum deprivation with additional thymidine block. This suggests that different synchronization techniques may potentially result in distinct gene temporal expression profiles. In contrast, patterns of temporal transcription of cell cycle-associated genes are similar to results of gene expression provided by the serum deprivation dependent techniques of Bar-Joseph, *et al.*, 2008 (Figure 2.21) [98].

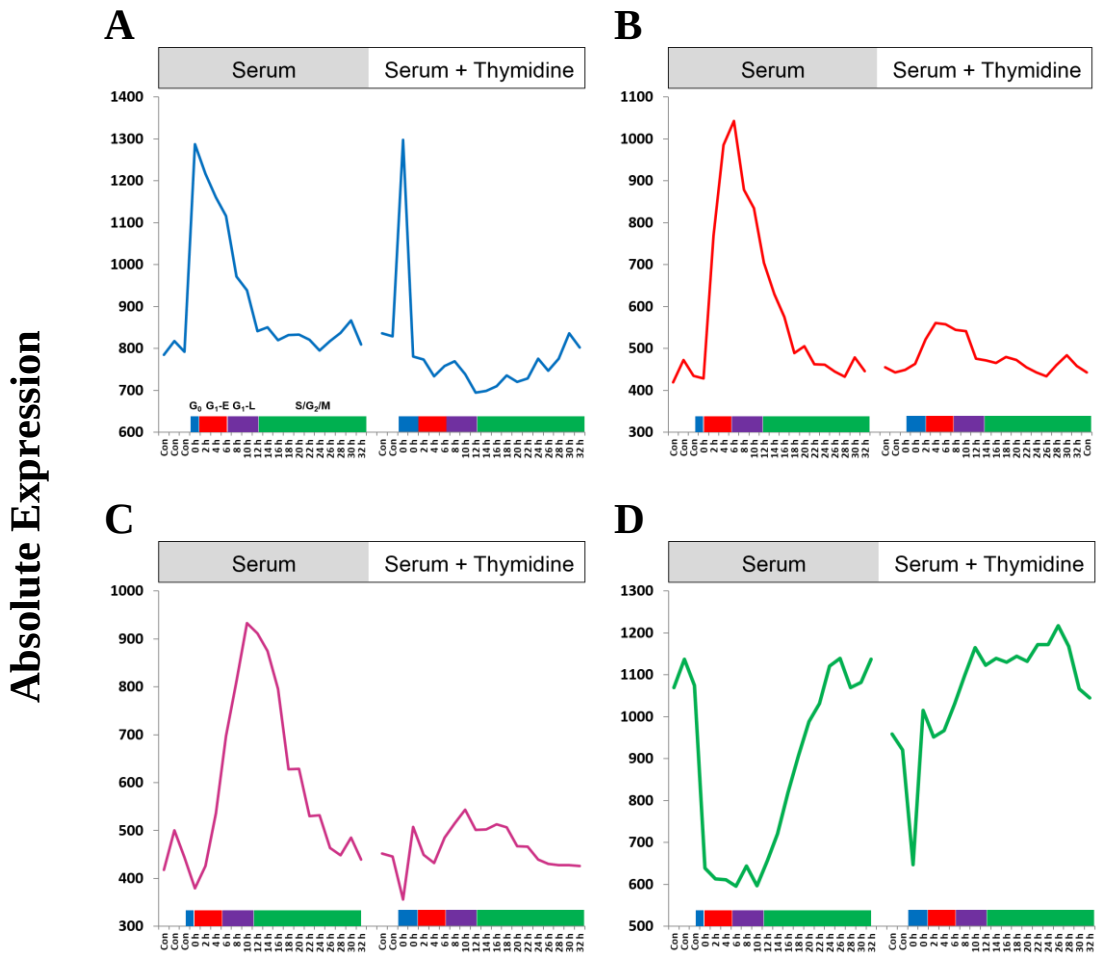


Figure 2.21: Unrepeatable expression features of genes. Gene expression profile of the cell cycle between fibroblasts synchronized by serum deprivation (plots with grey back ground) and fibroblasts synchronized by serum deprivation with additional thymidine block (plots in white background) across different time points. **A.** Peak expression at 0 h of genes after medium replacement. **B**. Peak expression between 0 and 6 h of genes after medium replacement. **C**. Peak expression between 6 and 12 h of genes after medium replacement. **D**. Peak expression of genes 12 h after medium replacement. Array dataset is courtesy of Bar-Joseph, *et al.*, 2008 [98].

Discussion

In order to identify cell cycle-associated genes up-regulated in S, and G_2/M-phases, human NHDF cells were synchronized by serum deprivation and RNA was extracted for microarray analysis followed by network analysis. Results demonstrate that serum deprivation was successfully used to generate a major population of cells with a synchronized cell cycle. The effectiveness of cell cycle synchronization was determined by flow cytometry. Under the condition of starvation, the cell cycle regulation system holds cells at a quiescent stage (G_0) which is a natural physiological response to serum deprivation [260].

To understand the cell cycle regulation mechanism at the transcriptional level, two array experiments were used to globally survey the pattern of gene expression. Genes with a similar profile of temporal expression were clustered by network structure analysis. The presence of known cell cycle genes in four major clusters was used to identify which phase of the cell cycle that each cluster was associated with. To further understand the biological role of clustered genes, the enrichment of GO term analysis was used to functionally profile the set of genes in a cluster. Four major clusters of genes associated with the cell cycle were found to be enriched for the terms associated with 'sterol metabolic process (G_0)', 'transcriptional regulation (G_1-Early)', 'protein transport (G_1-Late)', 'DNA replication (S)', and 'mitotic sister chromatid segregation (G_2/M)'. This suggests that the lists of genes in these four major clusters are associated with the terms of particular biological processes in the cell cycle.

In order to profile the genes up-regulated in S-, G_2-, and M-phase, they were manually annotated using published data. Of the genes identified in the S, G_2/M-associated clusters, 484 genes were genes were already known to be involved in the cell cycle, 69 were novel genes with unknown function, 78 were genes with putative cell cycle function, and 75 were genes with a known function that is unrelated to the cell cycle. This suggests that clustering based on similarity of expression is a good way to study complex biological systems. Furthermore, using the latest array technologies, graphic based network structure analysis is capable of providing a new insight into cell cycle-associated gene profiling.

Synchronization of cells by serum deprivation

In order to survey the cell cycle-associated genes that peak in expression during S-, G_2-, and M-phases, it is important to first generate a major population of cells with at least one

synchronized cell cycle [7]. Synchronization by serum deprivation was attempted for three different cell lines; HEK293T, THP-1, and NHDF.

Synchronization of cell cycle was not achieved with HEK293T cells

The HEK293T cells were starved for 48 h and then stimulated by serum refeeding. The synchrony was then determined by flow cytometry 20 h after serum refeeding. The serum deprivation didn't arrest the cell cycle of HEK293T cells. This suggests that a serum-dependent technique is not appropriate to generate a major population of synchronized HEK293T. This is because the HEK293T is a cell line transformed with SV40 large T antigen. This is a proto-oncogene derived from polyoma virus SV40, which can potentially transform host cells into tumours. SV40 large T-antigen binds and inactivates tumour suppressor proteins (p53, p105). This causes the cells to leave G_1-phase and enter into S-phase, promoting DNA replication. HEK293T cells are therefore not good candidates for synchronizing by serum starvation since the presence of SV40 T antigen is promoting DNA replication [202].

Synchronization and loss of synchrony of THP-1 cells after first division

THP-1 is a suspension cell line cultured from the blood of an acute monocytic leukemia patient [9]. THP-1 cells were starved in 0.2% serum for 48 h and then stimulated by serum refeeding. Around 75% of the cell population was found in G_0/G_1-phase just 48 h after starvation suggesting that serum deprivation can arrest the cell cycle of THP-1 cells. However, only 18% of cells left quiescent stage and re-entered the cell cycle 24 h after serum refeeding. It has been suggested that the synchrony of cells diminishes each time they pass the G_1-phase in an exponential culturing system i.e. each cell divides independently with unsynchronized cell cycle in the exponential growth culturing system [7]. The longer the doubling time the higher the possibility that cells may lose their synchrony between each other. It might be that a long doubling time for THP-1 cell proliferation (the average doubling time is 50 to 70 h [9]) results in a smaller synchronous cell population than cell types with a shorter doubling time, such as NHDF, which has a doubling time of about 25 hours [8]. Nevertheless, it has been suggested that an ideal cell line for synchronization should not have doubling time above 30 h [9]. Serum deprivation is therefore not believed to be the best way to synchronize the cells with long doubling time.

Serum deprivation can generate a major population of synchronized NHDF cells

The serum dependent technique can efficiently generate a major group of NHDF cell population at a quiescent stage; 85.2% at G_0/G_1-phase, 8% at S phase, and 6.9% at G_2/M-phase (Figures 2.5 and 2.6). The NHDF cells were released from the quiescent stage reaching mitosis 18 h after serum refeeding (Figure 2.6). This renders a major population of cells with synchronized transcriptional activity and is consistent with the result of array analyses [22,96,98]. Encouraged by the effectiveness of serum deprivation, NHDF cells were therefore chosen to be synchronized and to be used for array analyses.

This technique has been widely used in a number of important microarray experiments to classify the cell cycle genes [22,96,98]. It is also has been suggested that the feature of transcriptional regulation is relevant to the cell physiology in cells synchronized by serum starvation [261]. In contrast, the serum independent technique, such as chemical synchronization, has disadvantages in synchronizing the cells. It is believed that chemical treatment is potentially linked to the possible disruption of normal cell cycle regulation unbalancing the cell growth. For example, Gong *et al.*, 1995 [262] showed that cyclin E and B1 were over expressed in thymidine, mimosine, or aphidicolin synchronized human cells. In addition, Kung *et al.*, 1990 [263] suggested that short term exposure of aphidicolin or colcemid between 12 to 18 h may result in an unscheduled G_2/M-phase in the cell cycle. To study the cell cycle, it is therefore important that serum deprivation can provide cells with synchronized transcriptional activity at normal expression level.

In order to check the synchrony of the cell cycle after serum refeeding, total RNA was isolated from NHDF cells at a sequence of time points with regular intervals and then analysed by RT-qPCR. For marker up-regulated in S-, G_2-, and M-phase, the profile of mRNA expression suggests that the G_2/M-phase genes *CDK1* and *BUB1* were both up-regulated at 24 h following the serum refeeding. This confirmed the achievement of synchrony and was consistent with the results obtained by flow cytometry (Figure 2.7).

Reproducibility of biological replicates of array samples

A Pearson correlation matrix of the normalized signal intensity of array samples were calculated to test whether the two experiments were reproducible. First, biological repeats of array samples in both experiments were placed together (Figure 2.9). Second, the position of array samples in the clustering hierarchy tended to follow the time course (Figures 2.9B, C, E, F, G, H, I, J). It can be concluded that the two array experiments were reproducible.

Batch effect between experiments

After the reproducibility had been validated by a Pearson correlation matrix, network structure analysis was used to cluster genes based on their temporal expression profiles. It was found that of two major clusters, the level of batch effect expression profile was higher in one experiment than another and vice versa (Figures 2.13A and B). To further test whether the associated terms of biological process of identified genes were not associated with the cell cycle, GO term enrichment analysis was used to profile the list of genes. The result suggests that the associated terms of biological process are not associated with the cell cycle, such as 'Parkinson's disease'. It is possible that these genes were randomly expressed.

GO slim mapping of cell cycle-associated genes

In order to give a quick and broad profiling of the distribution of genes in the cell cycle, the list of cell cycle-associated genes were mapped with GO slim, a cut-down version of the GO ontologies. GO slim consists of a subset of terms compared to the full version of GO. The number of G_0-phase genes, which are associated with the terms of specific biological events is higher than the number of genes belong to S-, G_2-, and M-phase e.g. signal transduction, multicellular organismal development, and metabolic process. In contrast, the list of gene relevant with the term for cell cycle for S-, G_2-, and M-phase is higher than for G_0-phase. Relevant biological processes with the associated term of cell cycle of the listed genes in S-, G_2-, and M-phase suggests that the cells were entering the mitosis.

GO term enrichment analysis of cell cycle-associated genes clustered by network structure analysis

Following a quick GO slim profiling of genes identified by network structure analysis, the full-scale GO analysis was used to further subset the genes by associated terms of biological process. In the major cluster associated with G_0-phase, around 1780 genes up-regulated during this phase were clustered into 28 clusters by network structure analysis (Figure 2.18). The lists of genes in the biggest cluster (688 genes) and third biggest cluster (120 genes) were enriched by GO analysis for the term associated with 'chromosome organization'.

A good example is that known gene i.e. inhibitor of growth family, member 4 (*ING4*) and tumour protein P53 (*TP53*) were both clustered together. The protein expressed by gene *ING4* involved in chromatin remodelling and TP53 regulation [264]. TP53 is a well-known tumour suppressor protein expressed at a low level in normal cells, involved in transcriptional activation, and DNA binding. In response to cellular stress, such as serum deprivation, the half-life of TP53 is dramatically increased resulting in fast accumulation of gene products in the cell arresting the cell cycle [265].

On the other hand, the list of G_0-phase associated genes in the second largest cluster (cluster 05, 487 genes) was enriched for the term 'lipid biosynthetic processes. It has been reported that serum deprivation dramatically increases the level of lipid biosynthetic process in order to increase lipid stability in stressed human cells for survival [266]. The up-regulation of genes associated with the biological process of lipid biosynthesis suggested that cells were under stress caused by serum deprivation.

G_0-phase genes up-regulated in fibroblasts were previously reported by Coller *et al.*, 2000. The author suggested that: i) quiescent stage is not an extended period belonging to G_1-phase, but an individual phase with correlated regulatory genes based on identified expression features from array analysis, ii) genes with up regulation during this stage suggests that quiescence is an active stage of the cell cycle, and iii) many changes of gene expression levels were found after arresting the cell cycle for more than a 24 h arrest [22]. In addition, the quiescent-associated genes were found to be involved in the suppression of apoptosis and differentiation, which prevented the reversible cell cycle arrest. This explains that serum refeeding cannot stimulate entire population of starved NHDF cells [98,267].

There are 27 clusters (885 genes) and 17 clusters (881 genes) found to be associated with the early and late stage of G_1-phase in the cell cycle, respectively (Figures 2.18 and 2.19). In the early stage of G_1-phase, the list of genes in the biggest cluster (724 genes) of the G_1-early phase major cluster was enriched for the term 'RNA processing'. In addition, the list of genes placed in the second biggest cluster (132 genes) of G_1-early phase major cluster was enriched for the term 'regulation of transcription'.

It has been suggested that due to the degradation of RNA molecules in G_0-phase, the proteins are therefore not able to resynthesize rapidly maintaining the balance in need. The efficiency of macromolecular syntheses is therefore only around one-third compared to regular cells [268]. Following serum refeeding, cells responded to the serum stimulation. In order to re-enter the cell cycle, the efficiency of RNA processing and protein re-synthesis must be recovered to regular efficiency in normal cells depending on high regulation of associated genes. The associated biological terms of 'RNA processing' and 'regulation of transcription' of genes up-regulated in the early stage of G_1-phase suggests that the cells were entering the cell cycle.

In the late stage of G_1-phase, the list of genes was found to be enriched for the terms associated term 'RNA processing' and 'intracellular transport', suggesting that the response of cells shifted from the transcriptional level to proteomic level. This is consistent with the discovery by Campisi *et al.,* 1984, that the cells initiated DNA synthesis 12 h after serum refeeding [260,269]. The change of biological activity starting from the early to late stage of G_1-phase suggests that i) cells re-enter the cell cycle, ii) cells started to prepare for the mitotic activity.

There are in total 9 clusters (982 genes) which contain identified genes up-regulated during S-, G_2-, and M-phase (Figure 2.21). In the GO enrichment, the list of genes placed in the biggest cluster of S-, G_2-, and M-phase (901 genes) was enriched for the terms 'cell cycle', 'M phase', 'DNA metabolic process', 'chromosome segregation', 'microtubule-based process', 'chromosome organization', 'mitotic cell cycle', and 'cell cycle checkpoint'. It is notable that the associated biological process of GO term 'cell cycle' was only enriched for the clusters of gene up-regulated in S-, G_2-, and M-phase. This suggests that genes with unknown function clustered with key cell cycle-regulated genes potentially involved in the progression of cell cycle. This result of GO term enrichment analysis suggests that using network structure analysis is a robust strategy to forecast the role of genes with unknown function based on the clustering of similar transcriptional features.

How to develop a functional list of cell cycle-associated genes

Due to the complexity of various annotations from identified genes, the functional gene list should be manually annotated. A functional list should include the stage which genes were associated with, function of expressed genes, and annotated by universal annotation system, such as cell cycle genes list of Coller, *et al.*, 2006 [22] and Bar-Joseph, *et al.*, 2008 [98]. In addition, different annotations of each gene should also be summarized in the table. For example, this list should include the symbol of genes, description of genes, genomic location of genes, associated phase in the cell cycle of genes, associated function of genes in the cell cycle e.g. Entrez ID [270], ENSEMBL ID [216], GO term annotation [271], UNIGENE ID [272], UniPort (universal protein resource) ID which catalogue functional proteins supporting biological research [273], and reference numbers of genes with associated functions.

Functional annotation of genes manually curated according to published literatures

Given a functional gene list curated mainly based on the literature mining, we found that 484 of 706 genes of interest were involved in important mitotic cell cycle biological processes, including DNA repair, synthesis, chromatin organisation, and centrosome duplication in S-phase. The M-phase biological processes include cytokinesis. This key evidence of gene expression profile of cell cycle and other clustered genes sharing similar periodicity strongly suggesting these groups of genes are relevant to the cell cycle.

Our analyses based on the manual curation found 78 co-expressed genes of interest were putatively involved in cell cycle-associated events but without clear evidence according to literature mining. 69 co-expressed genes of interested were with unknown function. To forecast their putative role in the cell cycle regulation, subcellular localisation of proteins encoded by the genes of interest were curated according to literature mining. It is interesting that they are mostly located in association with the nucleus, suggesting that there may be numerous unknown pathways involving this cell cycle regulatory network.

In contrast, for the co-expressed genes labelled as unknown cell cycle gene (i.e. 69 genes), there was a lack of information to provide their location in the cells according to literature mining, suggesting less than half are associated with the nucleus (i.e. 22 genes). The others have little information on their location and associated biological events in the cell cycle.

Further experimental analyses are needed to identify and characterise their role in the cell cycle.

In comparison with the GO analyses that enrich the major biological events of associated genes, functional annotation evidenced by published literature providing detailed information from major biological processes to subcellular location in our analysis.

The transcription feature in cells synchronized by the serum dependent technique is different from chemical treatment

It has been suggested by Coller *et al.*, 2006 that the population of genes up-regulated in G_0-phase may potentially be different due to the techniques used to arrest the cell cycle [22]. This suggests that different conditions of synchronization may result in various transcription profiles in the cells. However, they only provided evidence at the quiescent stage rather than an entire cell cycle. Therefore, to further profile the difference of gene expression between NHDF cells synchronized either by chemical or serum dependent techniques, the data set of Bar-Joseph, *et al.*, 2008 was reanalysed by network structure analysis [98]. In that experiment, chemical and serum dependent techniques were both used to synchronize cells and then profiled transcription by array analyses. Time series dataset of gene expression were then clustered into four groups of cell cycle-associated genes.

In Figure 2.15, the transcription profile of cell cycle-associated clusters suggests that the temporal expression level of genes in cells synchronized by serum deprivation were up-regulated following a complete cell cycle. It was also found that the profile of temporal gene expression in NHDF cells synchronized by serum deprivation shared higher similarity between my results and those of Bar-Joseph, *et al.*, 2008 suggesting the reproducibility between two individual experiments [98]. In contrast, cells synchronized by thymidine block didn't provide a cell cycle expression profile associated with each phase of the cell cycle. This evidence may explain why chemical dependent synchronization may potentially induce the transcription activities irrelevant to the cell physiology [261-263].

In summary, a total number of 706 genes were thought to be associated with cell proliferation. Of this cell cycle list, 69 genes were co-expressed during S-, G_2-, and M-phase. To further investigate their role in cell division, loss of function analysis (i.e. RNAi), and subcellular protein localisation were used to characterise their role in the cell proliferation in following chapters i.e. chapter 3 and 4.

Chapter 3.

RNA interference targeting of novel cell cycle-associated genes

Introduction

In Chapter 2, we identified 706 genes up-regulated during the S-, G_2- and M-phase, 69 have not previously been associated with the cell cycle and potentially represent new components of the cell cycle machinery. In order to provide evidence that these unknown cell cycle-associated genes are involved in cell proliferation, we developed an esiRNA screening assay to identify the effect of gene knock down on the cell proliferation by using a RTCA system [115].

RNAi induced gene silencing is a mechanism that mediates mRNA degradation. Degradation is initiated by the presence of double-stranded RNA (dsRNA) that shares homologous sequence with a targeted gene [121]. However, genes sharing similar sequence to a targeted gene may result in an off-target effect of RNAi induced gene silencing [140].

To minimize the off-target effects, Ralf *et al.,* 2007 [132] and Frank *et al.,* 2006 [144] developed esiRNA. This process of enzymatically preparing dsRNA provides a heterogeneous population of siRNAs that target many different regions of an mRNA, and results in a dilution of off-target effects. esiRNA has been shown to efficiently knock down the expression of targeted genes in human cells [145]. Since gene silencing was first demonstrated in mammalian cells by Elbashir *et al.,* 2001 [121], this technology has been widely used to investigate the function of genes in biological systems combined with HTS [125].

The application of RNAi screening combined with HTS has been used to investigate gene function in cells from many different species, for example *Drosophila* [130,139], *C. elegans* [131], mouse [133], and human [132,146]. In order to deal with the large number of measurements from silenced genes in HTS, most researchers use a Z-score to determine the hits from a screen [130-135,147,148]. A Z-score is a statistical measure of a score's relationship to the mean in a group of scores quantifying the effect of RNAi on cell proliferation.

This type of analysis calculates a Z-score for a specific time point. However, end-point analysis can only provide a 'snapshot' at a particular time point and may miss differences that might be present at other times during the experiment. In order to analyse the real time data from the RTCA experiments, it is therefore necessary to calculate a score based upon all or part of RTCA profiles.

The best way to analyse the effect of RNAi on the cell proliferation is therefore over time to monitor the effect of gene knock down. RTCA is a technology that indirectly monitors cell proliferation based on changes in electrical impedance as cells divide in a culture dish. Electrical impedance of cells is detected via a gold microelectrode array that covers the base of a well and recorded as CI, based on the electrical conductance of the membrane lipid bilayer of living cells. [145,154-157]. In an RTCA experiment, CI increases as cells grow; this is roughly equivalent to cell proliferation although CI also increase if cells spread out and adhere more tightly to the base of the well. We have used RTCA analysis to analyse the effect of knocking down specific proteins and used a modified Z-score cut-off suggested by *Zhang* et al., 2011 to quantify the effect of RNAi on the cell proliferation [115,274].

In this analysis, 42 and 28 genes with unknown and known function in the cell cycle were screened by RNAi. In doing so, I have provided supporting evidence that these genes are novel components of the cell proliferation. I have carried out RNAi induced gene silencing experiments to demonstrate the effect of their knock down on the proliferation profile of cells as monitored by RTCA system.

Materials and methods

The RTCA system is a highly sensitive real time cell monitoring system that utilizes an array of electronic sensors on the base of cell culture wells to measure impedance. In order to optimise conditions for the RNAi screen, a number of parameters were investigated. Firstly, laminin coating the wells was tested to evaluate its effect on cell proliferation compared with uncoated wells. Secondly, in order to analyse the effect of RNAi on CIGR over time, an experiment was performed to find the optimal seeding density. Thirdly, knowing that cytotoxicity of the transfection reagent can lead to a synchronized cell cycle of plated cells [115], a number of transfection reagents were tested to evaluate their effects on cell proliferation. Fourthly, the knock down efficiency using two types of RNAi was compared and contrasted (Smart Pools of siRNA vs. esiRNA). Unknown cell cycle genes identified in Chapter 2 were then screened for the effect of RNAi knockdown. The knocking down that decreased or increased cell proliferation, was determined by a process of evaluation and scoring.

RTCA measurement of cell proliferation

The real time cell analyser (RTCA, xCELLigence Roche, Penzberg, Germany) system was used to monitor the change of cell impedance over time. A RTCA profile of cell dynamics is positively correlated with the coverage of adherent cells and detected by microelectrodes in the base of wells [275]. Prior to measuring CI before and after transfection across a number of time points, a background impedance measurement was recorded by adding 100 µl of DMEM with 10% (v/v) FBS and with antibiotics (25 U/ml penicillin and 25 µg/ml streptomycin) to all wells of the plate. Cells at seeding density of 6,000 cells/cm^2 were then plated equally in each well with 100 µl of complete medium.

The 96 well E-plate (with NHDF cells) was then docked on the xCELLigence cell analyser SP (96-well signal E-plate) station and recording of CI commenced (Figure 3.1 A). Readings were taken every 15 min for the first 200 sweeps (turns), every 30 min for the next 200 sweeps, and finally at 60 min intervals for the last 100 sweeps. Data was obtained as an Excel file with a column for each well and a row for each time point.

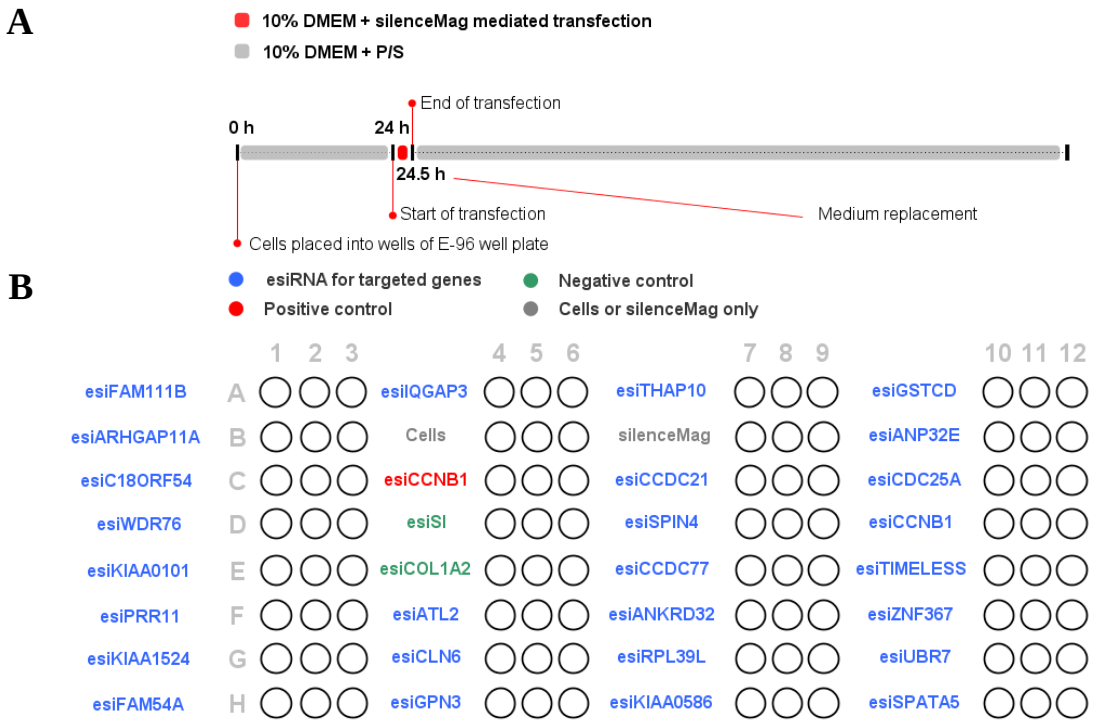


Figure 3.1: Experimental design of 96 well E-plate for real time RNAi screening. A) Time line of experiment from cell plating at t = 0 h to transfection conducted by silenceMag^{TM} at 24 h for 30 min followed by replacement of medium. The development of RNAi on cell proliferation was then monitored by RTCA system. **B)** Example of the layout of cells in a 96 well E-plate given esiRNA targeting the genes of interest.

Optimization of seeding density and plate coating

NHDF cells were plated in 96 well E-plates for the RTCA system at a seeding density of 250, 500, 750, 1,500, 3,000, and 6,000 cells/cm^2 in DMEM (Sigma-Aldrich, Gillingham, UK) with 10% (v/v) FBS (Heat inactivated serum, GE Healthcare, Little Chalfont, UK), and antibiotics (25 U/ml penicillin and 25 µg/ml streptomycin, Life technologies, Paisley, UK). Serially diluted cell densities were plated to test the correlation between seeding density and the exponential growth phase in cell proliferation. Cells were grown in 5% CO_2, at 37^{o}C.

To optimize conditions for culturing NHDF cells in 96 well E-plates, laminin (Life technologies) at concentration of 50 µg/ml was used to coat wells and was compared to uncoated wells. Laminin was added to each well and left for 3 h, before washing with 1xPBS three times.

Optimisation of Transfection

Lipofectamine® 2000 transfection

Lipofectamine® 2000 (Life technologies) is a cationic lipid compound mediating the formation of RNA-cationic lipid complexes which fuse with the negative charged cell membrane [276,277]. NHDF cells were plated in 6 well plates in complete medium at a seeding density of 6,000 cells/cm^2 and cultured for 24 h before addition of transfection reagents. For the transfection of each well, esiRNA was prepared by combining 3 µl of esiRNA at a stock concentration of 0.5 or 5 µM (for the final concentration at 5 or 50 nM of esiRNA, respectively) and 27 µl Opti-MEM (Life technologies). To prepare the transfection reagent, 1.5 µl of lipofectamine® 2000 was mixed with 28.5 µl Opti-MEM. The solutions were then left at room temperature for 5 min, before being combined and left at room temperature for a further 25 min (total volume 60 µl). The combined solution was then mixed with 240 µl serum-free medium and added to NHDF cells. Cells were cultured for 24 h in an incubator at 37ºC with 5% CO_2 before being replaced with complete medium.

SilenceMag™

The transfection reagent silenceMag™ (OZ bioscience, Marseille, France) comprises of positive nanomagnetic particles mediating the formation of RNA-particle complexes, delivering RNA by the magnetic force provided by a strong magnetic plate (OZ bioscience) [278]. For each sample of silenceMag™ reagent transfections, 6 µl of esiRNA at 5 µM or 0.5 µM (for the final concentration of esiRNA at 5 or 50 nM respectively) was mixed with 6 µl ddH_2O and 6.6 µl reagent silenceMag™. The transfection reagent mix was added to a total volume of 600 µl antibiotic free medium which was then added to each well at the seeding density of 3,000 cells/cm^2. Cells in a 6 well plate were placed on a strong magnetic plate in an incubator at 37ºC with 5% CO_2 for 30 min before replacing the medium with complete medium. Samples of total RNA were harvested from each well individually.

RT-qPCR

To test the knock down efficiency of targeted genes selected by array dataset analysis, cells were analysed at 48 h after esiRNA induced gene silencing and knock down tested by RT-qPCR. RNA samples were isolated by adding 350 µl buffer RLT (Lysis buffer for lysing cells prior to RNA isolation, QIAGEN, Manchester, UK) per well of a 6 well plate at 48 h after transfection. Extraction was performed using the RNeasy Mini Kit (QIAGEN) and total RNA was isolated according to manufactures' instruction. 0.5 µg of total RNA was used for RT-qPCR according using methods described on page 49. The sequences of primers used to

analyse known and unknown cell cycle genes are shown in Table 3.1. GAPDH was used as a control.

Analysis of cell viability

Before transfection, NHDF cells seeded at a density of 6,000 cells/cm^2 in 25 cm^2 TC flasks were cultured for 24 h in DMEM (Sigma-Aldrich) with 10% (v/v) FBS (GE Healthcare), and antibiotics (25 U/ml penicillin and 25 µg/ml streptomycin) (Life technologies). SilenceMag™ transfection with 50 nM esiRNA was performed in antibiotic-free DMEM with 10% (v/v) FBS for 30 min before replacing with complete medium. Samples were harvested 72 h after transfection and 1 × 10^6 cells were stained with SYTOX® Blue (Life technologies) at 1 µM for 5 min before analysis using BD FACS Aria lllµ Cell Sorter (BD bioscience, Oxford, UK). Data was analysed by BD FACSDiVa™ software (BD bioscience).

Gene	Sequence (5' to 3')
ARHGAP11A	F-GTTGGTTGGCGACTTGCAAA R-GTCCTGTCCAAGACATCCGG
ATL2	F-ACTCTGGGGAGTTCAGAGGAA R-TGGATGACAGTCTCTGTCGC
KIAA1549L	F-GACCTACCGCCCAGAAATGT R-CCGATGTCTGAGGTGGATCG
C17orf53	F-CCCAAACTCCAACCCATGGT R-GCGTGGATTTCATGGTCAGC
CCDC77	F-TGCCCGTTATGCATGAGAGT R-TTCCTCACGAATTCGGGCAA
CCNB1	F-TGACTTTGCTTTTGTGACTG R-GTGTCCATTCACCATTATCC
FIGNL1	F-GCTTGCACAGTGCTTGGTAG R-TATGCATCTGCCTTCGGTCC
FOXM1	F-GGAGGAAATGCCACACTTAGCG R-TAGGACTTCTTGGGTCTTGGGGTG
REEP4	F-TACATCGTGCAGGCCAAGG R-TCCAGGTAGAGGGGGTCATG
RIBC2	F-AGCAGATCCGCCTAGTCCA R-GAGGTTGCTGCTCCAGAC
SI	F-AGTTACCGGTGGCATTCTGG R-ATGCTGGCATTGCTGGTAGT
ZNF367	F-CTACGCCAGGCTGAAGAGAG R-GCTTGCCTTTCAAAGTGGGG
GAPDH	Hs_GAPDH_2_SG QuantiTect Primer Assay (200) (QIAGEN, Manchester, UK)

Table 3.1: Primer sequences used in qPCR analysis. F and R refer to forward strand and reverse strand primers. qPCR, quantitative polymerase chain reaction.

RNAi screening in RTCA measurement over time

The library of esiRNAs (endoribonuclease-prepared short interfering RNAs) employed here has been described elsewhere [146,279,280]. Cell cycle-regulated gene *CCNB1* was chosen as a positive control. *CCNB1* is essential to regulation of mitosis [243,281]. Two types of negative controls were used. Firstly, esiRNA was used to target the *SI* gene, which is specifically expressed in small intestinal enterocytes and not expressed in fibroblasts. Secondly, esiRNA was used to target *COL1A2*, a gene expressed in fibroblast but not thought to effect cell proliferation when expression is perturbed. In comparison to the gene *COL1A1*, which plays an essential role in fibril-forming collagen in most connective tissue, mutation of *COL1A2* seems to be less important in the formation of collagen fibrils [282]. To perform the screening, three wells of a 96-well plate were used for each esiRNA and each plate was repeated three times.

In this analysis, there were 21 96 well E-plates used to screen 42 unknown and 28 known cell cycle-associated genes classified from two typical array experiments (Figure 3.1 and Table 3.2). Unknown and known cell cycle-associated genes were selected according to the scoring of their expression profile developed by senior statistician Dr. Helen Brown at The Roslin Institute. A score was developed to provide an objective measure of the 'importance' of a gene. A range of features from gene expression profiles were constructed and assessed statistically in terms of their ability to predict whether the genes were classed 'Very good', 'Good' or 'Medium' following a visual assessment of the gene expression profiles by Tom Freeman. Ordinal logistic regression was used to determine the most predictive weighted sum of the selected features. The score was developed using data from existing datasets and was calculated as:

- Score = 10 - 7.553*Max_diff + 5.394*Max_ratio + 4.613*Max3 - 18.430*Slope_pos + 14.426*Slope_neg
- Max_diff = maximum difference from baseline (0-6 hour average) over profile
- Max_ratio = maximum ratio to baseline (0-6 hour average) over profile
- Max3 = maximum average of 3 adjacent measurements (taken at 2 h intervals) over profile
- Slope_pos = slope during 0-6 hour period, but set to zero if negative
- Slope_neg = slope during 0-6 hour period, but set to zero if positive

Each term was calculated from the natural logarithms of expression values plus 1.

Data transformation of cell impedance index

Calculation of CIGR was performed using the R package RTCA (version 1.12.0) using R software (version 2.14.2) [208,283]. Figure 3.2 and 3.3 demonstrates the growth profiles of positive and negative controls as CIGR over time. The RTCA package was used to transform CI over time into CIGR at regular 30 min time intervals during the course of experiment (Figure 3.2 B) [115,284]. Average of the CIGR over time per well was calculated as a Z-score representing the effect of RNAi (Figures 3.2 B and C).

An example of Z-score normalisation is shown in Figure 3.3. For the raw value (i.e. averaged CIGR) of each sample (Figure 3.3A), its biological and technical triplicates (Figure 3.3D) were normalised using the robust version of Z score normalisation (Figure 3.3B), minimising the variety between each sample (Figure 3.3C).

Quality assessment and scoring of collected dataset

Quality assessment of CI data set was performed in R [285]. To evaluate the quality of the dataset, quantile-quantile plots (q-q plots) were used to compare the distribution between the collection of data and theoretical distributions [286]. The degree of non-parametric approach between normal theoretical quantiles and collected data quantiles determined 'the goodness of fit'. The higher the better goodness of fit between the collected data and theoretical distribution, the better the quality of the collected data.

A	S-, G_2-, and M-phase genes with unknown function (42)	Ensemble No.
ANKRD32	Ankyrin repeat domain 32	ENSG00000133302
ANP32E	Acidic (leucine-rich) nuclear phosphoprotein 32 family, member E	ENSG00000143401
ARHGAP11A	Rho GTPase activating protein 11A	ENSG00000198826
ARHGAP11B	Rho GTPase activating protein 11B	ENSG00000187951
ATL2	Atlastin GTPase 2	ENSG00000119787
C17orf53	Chromosome 17 open reading frame 53	ENSG00000125319
C18orf54	Chromosome 18 open reading frame 54	ENSG00000166845
C19orf48	Chromosome 19 open reading frame 48	ENSG00000167747
C4orf46	Chromosome 4 Open Reading Frame 46	ENSG00000205208
C5orf34	Chromosome 5 open reading frame 34	ENSG00000172244
CCDC15	Coiled-coil domain containing 15	ENSG00000149548
CCDC18	Coiled-coil domain containing 18	ENSG00000122483
CCDC77	Coiled-coil domain containing 77	ENSG00000120647
CLN6	Ceroid-lipofuscinosis, neuronal 6, late infantile, variant	ENSG00000128973
DEPDC1B	DEP domain containing 1B	ENSG00000035499
FAM111A	Family with sequence similarity 111, member A	ENSG00000166801
FAM111B	Family with sequence similarity 111, member B	ENSG00000189057
FAM161A	Family with sequence similarity 161, member A	ENSG00000170264
FAM72B	Family With Sequence Similarity 72, Member B	ENSG00000188610
FIGNL1	Fidgetin-like 1	ENSG00000132436
FOXRED1	FAD-dependent oxidoreductase domain containing 1	ENSG00000110074
GPN3	GPN-loop GTPase 3	ENSG00000111231
GSTCD	Glutathione S-transferase, C-terminal domain containing	ENSG00000138780
HIST2H4B	Histone cluster 2, H4b	ENSG00000270276
IQGAP3	IQ motif containing GTPase activating protein 3	ENSG00000183856
KIAA0586	KIAA0586	ENSG00000100578
KIAA1524	KIAA1524	ENSG00000163507
KIAA1549L	KIAA1549-like	ENSG00000110427
MGME1	Mitochondrial Genome Maintenance Exonuclease 1	ENSG00000125871
MTFR2	Mitochondrial Fission Regulator 2	ENSG00000146410
PRR11	Proline rich 11	ENSG00000068489
REEP4	Receptor accessory protein 4	ENSG00000168476
RIBC2	RIB43A domain with coiled-coils 2	ENSG00000128408
RPL39L	Ribosomal protein L39-like	ENSG00000163923
SPATA5	Spermatogenesis associated 5	ENSG00000145375
SPIN4	Spindlin family, member 4	ENSG00000186767
THAP10	THAP domain containing 10	ENSG00000129028
TMEM194A	Transmembrane protein 194A	ENSG00000166881
UBR7	Ubiquitin protein ligase E3 component n-recognin 7	ENSG00000012963
WDR76	WD repeat domain 76	ENSG00000092470
ZGRF1	Zinc Finger, GRF-Type Containing 1	ENSG00000138658
ZNF367	Zinc finger protein 367	ENSG00000165244

B	**Known genes (28)**	**Ensemble No.**
AURKB	Aurora Kinase B	ENSG00000178999
BUB1	BUB1 mitotic checkpoint serine/threonine kinase	ENSG00000169679
CCNA1	Cyclin A1	ENSG00000133101
CCNA2	Cyclin A2	ENSG00000145386
CCND1	Cyclin D1	ENSG00000110092
CCNE1	Cyclin E1	ENSG00000105173
CCNE2	Cyclin E2	ENSG00000175305
CDC20	Cell division cycle 20	ENSG00000117399
CDC25A	Cell division cycle 25A	ENSG00000164045
CDC25B	Cell division cycle 25B	ENSG00000101224
CDC45	Cell Division Cycle 45	ENSG00000093009
CDC6	Cell division cycle 6	ENSG00000094804
CDK1	Cyclin-Dependent Kinase 1	ENSG00000170312
CENPA	Centromere protein A	ENSG00000115163
CEP85	Centrosomal protein 85kDa	ENSG00000130695
E2F1	E2F transcription factor 1	ENSG00000101412
E2F7	E2F transcription factor 7	ENSG00000165891
E2F8	E2F transcription factor 8	ENSG00000129173
ESPL1	Extra spindle pole bodies homolog 1	ENSG00000135476
GINS1	GINS complex subunit 1	ENSG00000101003
KIAA0101	KIAA0101	ENSG00000166803
KIF11	Kinesin family member 11	ENSG00000138160
MAD2L1	MAD2 mitotic arrest deficient-like 1	ENSG00000164109
MCM10	Minichromosome maintenance complex component 10	ENSG00000065328
ORC1	Origin recognition complex, subunit 1	ENSG00000085840
PARPBP	PARP1 Binding Protein	ENSG00000185480
PCNA	Proliferating cell nuclear antigen	ENSG00000132646
SPACA1	Sperm acrosome associated 1	ENSG00000118434

C	**Controls (3)**		**Ensemble No.**
CCNB1	Pos	Cyclin B1	ENSG00000134057
COL1A2	Neg	Collagen, type I, alpha 2	ENSG00000164692
SI	Neg	Sucrase-isomaltase	ENSG00000090402

Table 3.2: Genes studied in the RNAi screening. List of genes includes **A)** genes with unknown function, **B)** known cell cycle-associated genes, and **C)** controls.

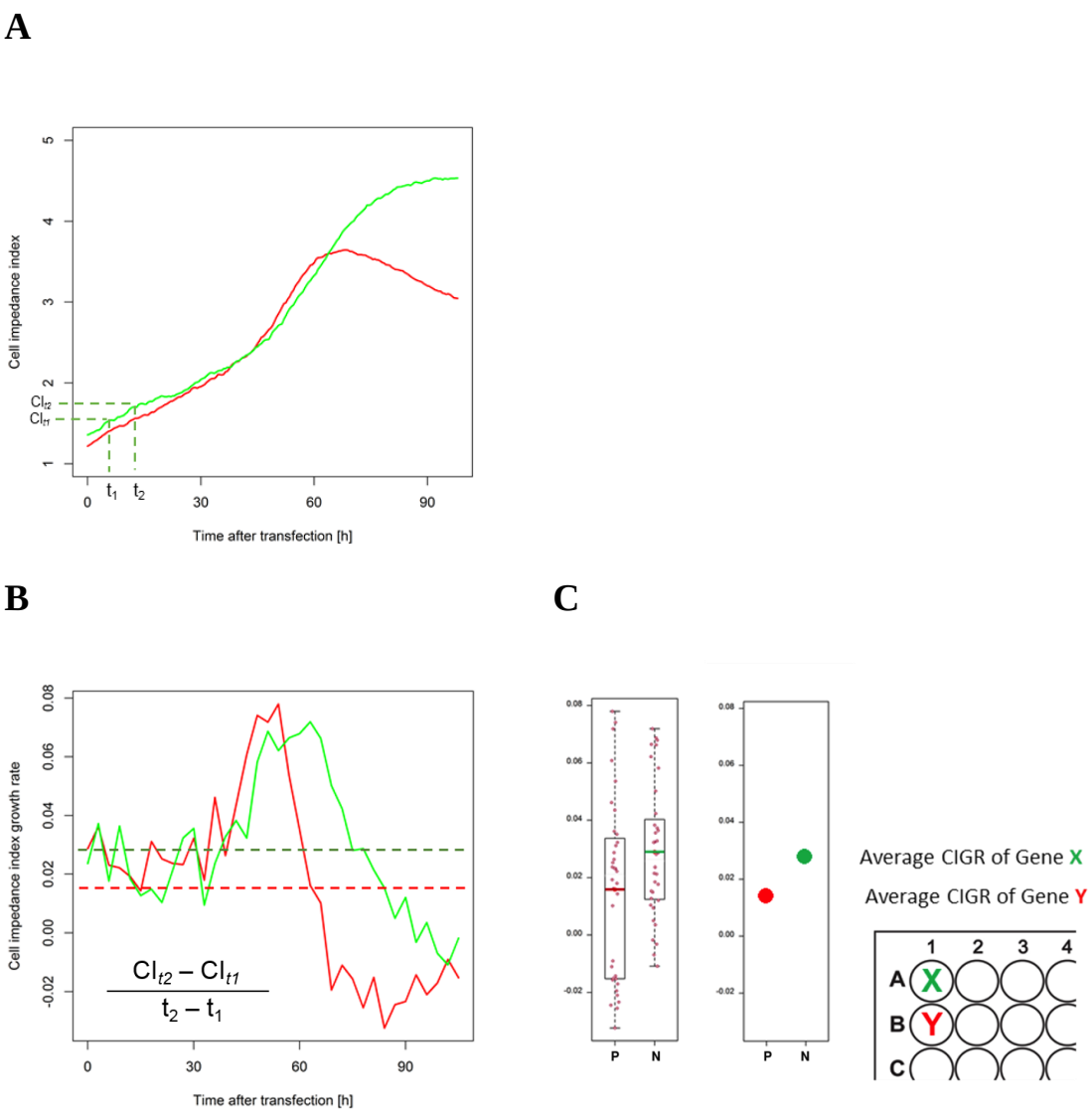


Figure 3.2: Example of the CIGR transformation and scoring of RNAi. A. Plots of CI over time after transfection. Red line represents cell proliferation affected by RNAi. Green line represents control. **B.** Transformation of CI to CIGR. Dash line in green as negative control and red as positive control calculated from the averaged of CI including each measurement time points. **C.** Boxplot of CI consists of a sequence of measurement time points after transfection. Green and red line is the average of CI as single number value represents the effect of RNAi after transfection in each well of a 96 well plate. **P** and **N** represent positive and negative control respectively.

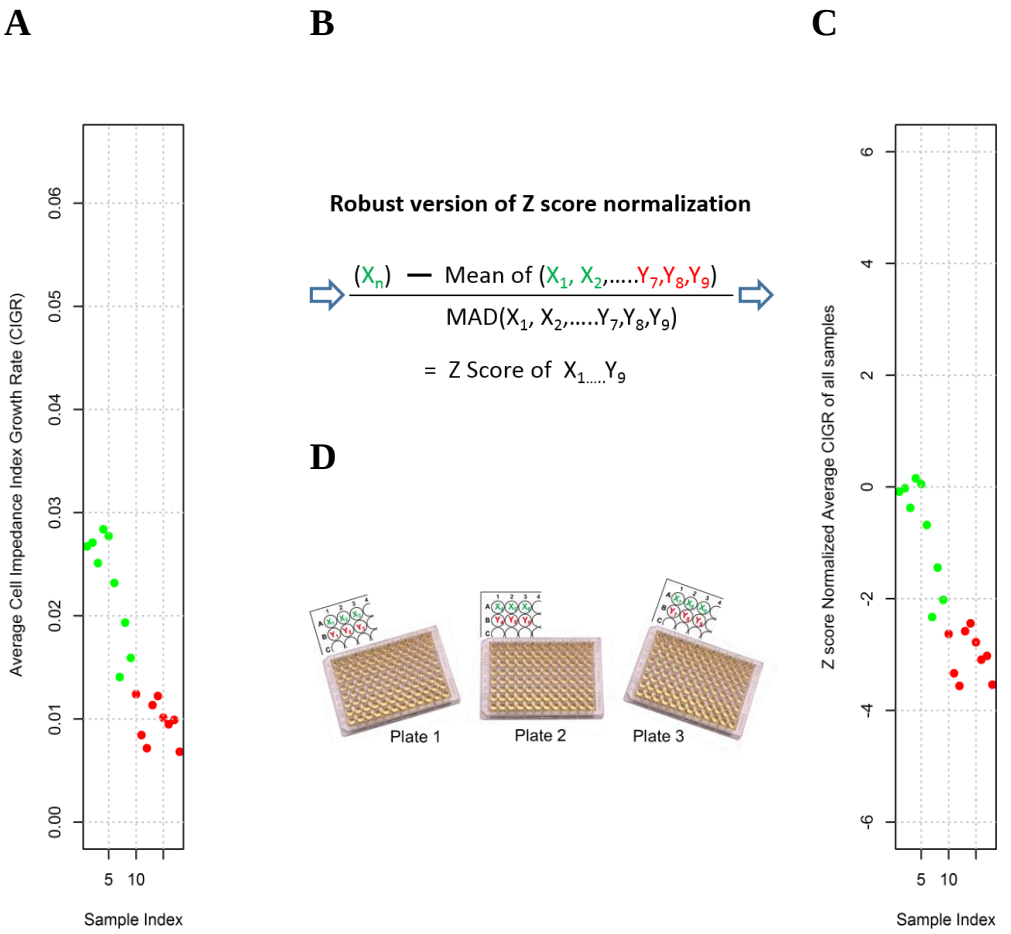


Figure 3.3: Examples of transformed CIGR values rescaled before and after the robust version of Z score normalisation. A) Raw value per sample of transformed CIGR from technical and biological triplicates i.e. n = 9 green/red spots. **B)** Equation used to normalise the raw value of transformed CIGR resulting in rescaled values. **C)** Normalised values after the robust version of Z score normalization. **D)** Examples demonstrate the wells in E 96-well plate related to technical and biological triplicates.

Score cut-off for RNAi screen

Z-score thresholding was performed in R using the RTCA package [283]. In order to compare the effect of esiRNA induced gene silencing and determine the hits, the robust version of Z-score based normalization of averaged CIGR was performed at regular time intervals prioritizing unknown novel genes for further analysis [274,287]. Robust version of Z-score normalization (robust Z-score = x – median (sample)/MAD (sample)*1.4826; MAD = median (|x – median (population of samples)|) is relatively tolerant to 'outliners' reducing their effect on the results (Figures 3.2 B and C). The score cut-off was determined by p-value estimated between samples and negative controls.

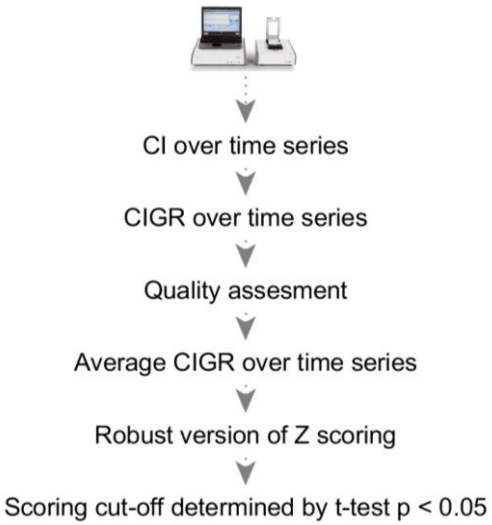


Figure 3.4: Flow chart of RTCA profile analysis determining hits of genes important in cell proliferation. Raw data of CI were extracted from RTCA system. Rational transformation of CI was then calculated as CIGR. Quality of dataset was then assessed by the degree of non-parametric approach between the distribution of collected data and theoretical normal distribution. Average of the CIGR was then calculated across each measurement time. Normalization was performed to reduce diversities between samples. Score cut-off of samples was determined depending on their p-value from t-test smaller than 0.05. Genes' role in dividing cells was then classified as activators or regulators by normalized scores.

Results

The RTCA system monitors the effect of RNAi on cell growth in real time. Firstly, to optimize transfection conditions, the seeding density and cell-plating environment were tested. Secondly, in order to find the best transfection techniques that minimized any side effects that might affect cell proliferation, silenceMag™, a non-cytotoxic magnet-assisted transfection reagent was tested in comparison with lipofectamine® 2000, a liposome-based reagent. In addition, to evaluate the knockdown efficiency provided by transfection reagents, RT-qPCR was performed to determine the effect of gene silencing. With diluted off-target effect, esiRNA was used to knock down genes that were identified in two array experiments. Using multiple statistical analyses, such as Z-score cut-off, an effect of RNAi on cell proliferation was identified with weak effect at 48 h and maximal effect at 72 h after transfection. Of the 42 unknown cell cycle-associated genes, 19 influenced the state of cell proliferation.

The exponential phase of growth was positively correlated with serially diluted seeding densities

In order to determine the optimal seeding density and the duration of the exponential growth phase in RTCA, serially diluted cells were plated on an E-plate 96. The length of the exponential phase of CI over time was positively correlated with seeding density (Figure 3.5).

The seeding of cells in preparation for transfection requires 24 h for adhering. The timing of maximal RNAi effect on cell proliferation varies between 24 and 72 h after transfection depending on which genes are targeted. Taken together, the minimal duration of cells at log phase is 96 h. The seeding density of NHDF cells at 6,000 cells/cm^2 results in stationary growth phase by 96 h, making it the optimal seeding density for in this experiment (Figure 3.5). It is interesting to note there was a rapid increase in the CIGR measurement between 0 and 2 h, which was due to the cells adhering to plate.

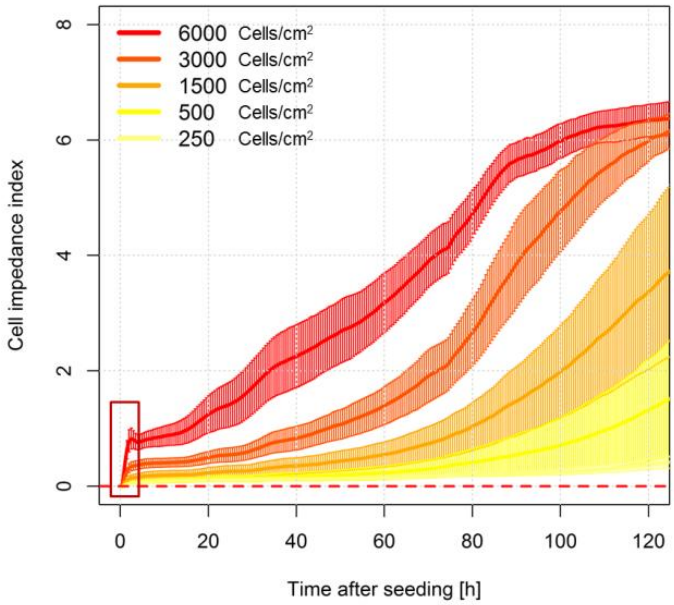


Figure 3.5: Duration of exponential phase was positively correlated with serially diluted seeding densities. The length of exponential phase in cell proliferation was positively correlated with serially diluted seeding densities of NHDF cells at 250, 500, 750, 1,500, 3,000, and 6,000 cells/cm^2 in E-plate 96. The CI over time of serially diluted seeding densities was profiled by RTCA.

NHDF cells can adhere to the wells of a culture dish without laminin coating

In order to find the best conditions for cell culture in RTCA system, a trace of CI over time in RTCA profile was used to determine the effect of laminin coating 96 well in E-plates. As can be seen in Figure 3.6, plated cells can adhere to plates with or without laminin coating. The trace of CI over time indicated cell proliferation was higher on wells without a laminin coating than wells with a laminin coating. The result suggests that laminin coating reduces the sensitivity of the sensor recording the CI.

Transfection reagent silenceMag^{TM} has minimal side effects on cell proliferation

In order to evaluate the potential side effects of transfection reagents on cell growth, two transfection reagents lipofectamine® 2000 and silenceMag^{TM} were compared to test for any inhibition in cell growth. Cell growth was arrested when the cells were treated with lipofectamine® 2000 until the replacement of complete medium (Figure 3.7A). The transfection reagent silenceMag^{TM} did not arrest the proliferation of cells (blue line) when compared to the CI over time of cell medium only. This suggests that treatment with silenceMag^{TM} had minimal side effects in cell proliferation (Figure 3.7B).

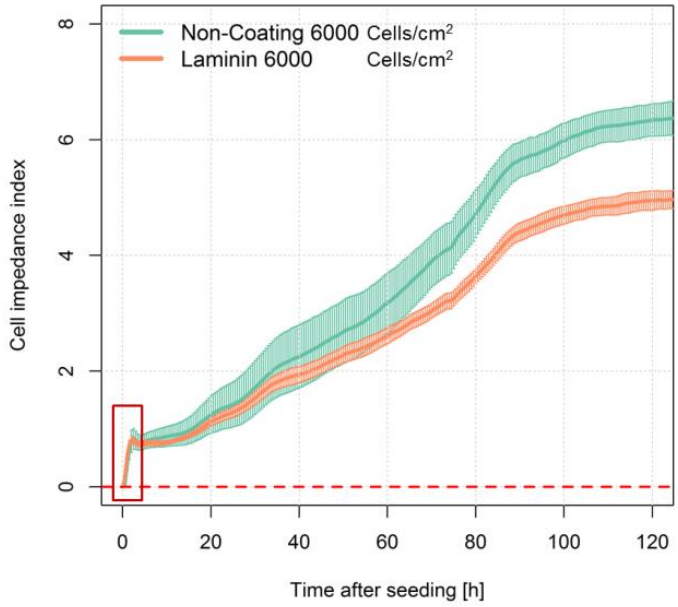


Figure 3.6: Laminin coating E-96 plate reduced the sensitivity of electrode sensor in RTCA profile. Trace of cell proliferation as CI over time in 96 well E-plate with and without laminin coating. Data shown as mean ± SD of triplicate wells.

A **B**

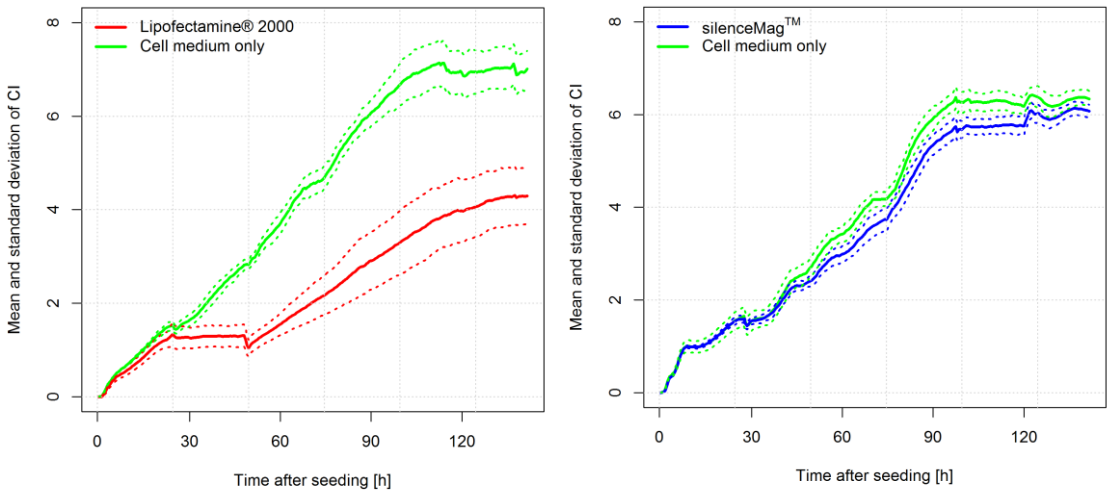


Figure 3.7: Transfection reagent silenceMag™ had minimal side effect in cell proliferation. A. The transfection reagent Lipofectamine® 2000 arrests the cell proliferation relative to untreated control (see Figure 3.7A red line) compared to cell medium only (green line). **B.** The transfection reagent silenceMag™ has minimal side effect on cell proliferation relative to untreated control. Dot line represents SD. Data shown as mean ± SD of triplicate wells.

Knock down efficiency of gene silencing

In order to find the best cell culture conditions for RNAi screening using the RTCA system, four combinations including two transfection reagents (Lipofectamine® 2000 and SilenceMag™) and two kinds of dsRNA (siRNA and esiRNA) were evaluated by testing knock-down efficiencies through RT-qPCR. The percentage knock-down was calculated based on the expression level of the housekeeping gene *GAPDH* (Figure 3.8).

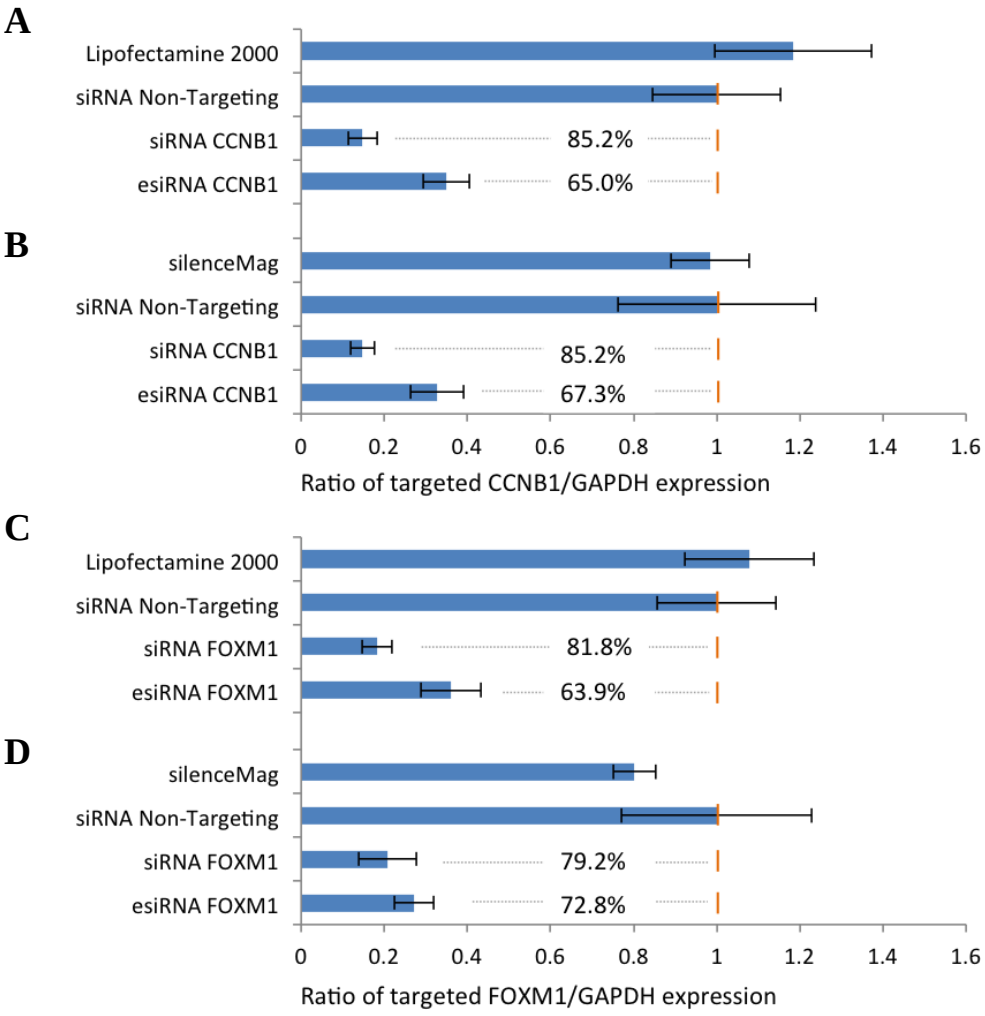


Figure 3.8: Knock down efficiency using two methods of RNAi was compared and contrasted (Smart Pools of siRNA vs esiRNA). A) Knock down efficiency of dsRNA or esiRNA CCNB1 induced gene silencing delivered by lipofectamine® 2000 conducted transfection. **B)** Knock down efficiency of dsRNA or esiRNA CCNB1 induced gene silencing delivered by silenceMag™ conducted transfection. **C)** Knock down efficiency of dsRNA or esiRNA FOXM1 induced gene silencing delivered by lipofectamine® 2000 conducted transfection. **D)** Knock down efficiency of dsRNA or esiRNA FOXM1 induced gene silencing delivered by silenceMag™ conducted transfection. Data shown as mean ± SD of replicates (n = 3). Expression ration was normalized in the unit siRNA Non-targeting (negative control).

Viability of NHDF cells after silenceMag™ conducted transfection

In order to ensure that the apparent reduction in cell proliferation was not due to cell death induced by the transfection reagents, the cell viability was measured. This was performed by flow cytometric analysis following staining the cells with SYTOX® Blue. The results in Figure 3.9 illustrate that transfection with esiRNA for the genes involved in cell proliferation (CCNB1 and FOXM1) did not affect cell viability when compared with the untreated control, or cells treated with esiRNA for COL1A2, a gene that is not involved in cell proliferation.

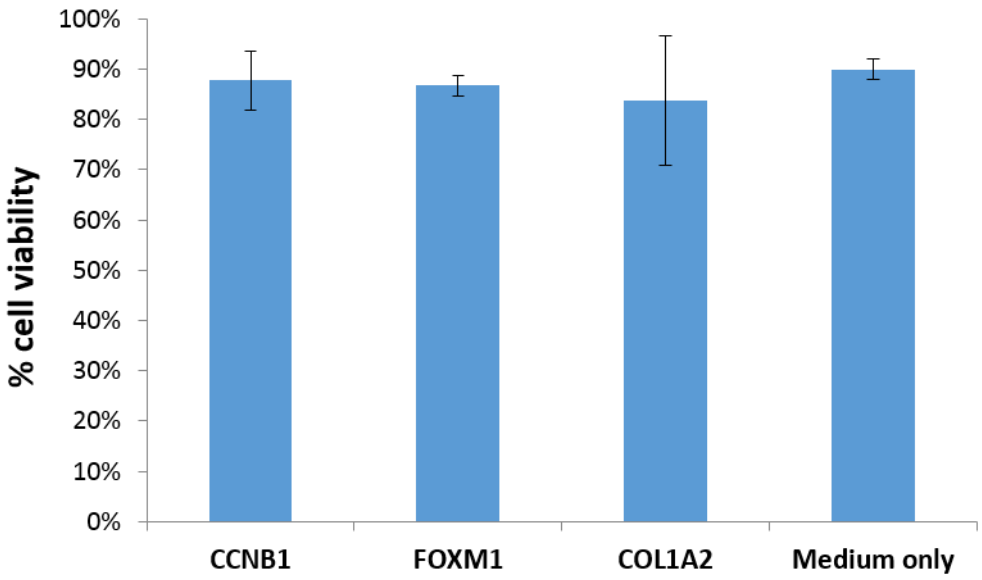


Figure 3.9: High viability of cells treated by esiRNA induced gene silencing 48 h after silenceMag™-conducted transfection. Results of cell viability analysis determined by flow cytometry using PI staining. Essential cell cycle-regulated genes *CCNB1*, *FOXM1*, and non-essential fibril-collagen gene *COL1A2* expressed in fibroblast were knocked down by esiRNA induced gene silencing. The population of live fibroblasts given esiRNA was determined by flow cytometry using SYTOX® Blue staining. The analyses were performed in three biological replicates for every esiRNA induced gene silencing. Error bars represent the standard deviation of replicates (n = 3). The cell concentration of the samples to be $1x10^6$ cells/ml.

Analysis of negative and positive controls of RNAi

An analysis was performed to investigative the effect of RNAi on cell proliferation using three genes selected as positive and negative controls (Figure 3.10). *CCNB1* is an important M-phase cyclin that conducts the biological processes in dividing cells. The CI trace (red) of cells given *CCNB1* esiRNA showed that cell proliferation began to decrease at 48 h after transfection (72 h after seeding). In contrast, silencing of the non-essential gene *COL1A2* or *SI* in the fibroblast, as negative controls, showed no effect on the CI over time (Figure 3.10 green lines). Differences between the negative and positive control genes were statistically significant 90 h after cell plating (transfection was performed at 24 h after cell plating). This suggests that esiRNA induced gene silencing potentially required more than 48 h to affect cell proliferation.

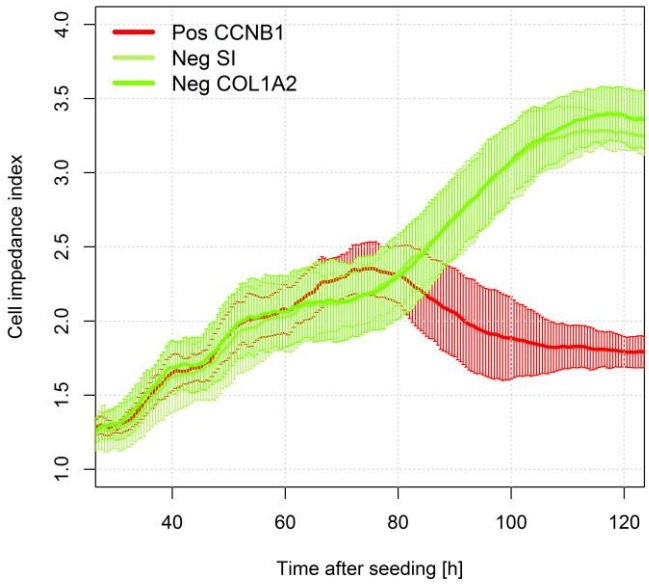


Figure 3.10: CI over time for the positive control and negative controls profiled by RTCA system. Monitoring of the CI over time of cells given gene silencing of the positive (gene *CCNB1*) and negative control (gene *SI* and *COL1A2*) were recorded by RTCA system after transfection.

Hits of genes in RNAi screening determined by Z-score cut-off

This experiment was designed to test the optimal conditions for monitoring gene silencing and cell division in real time using the RTAC system. There were 70 genes screened by RNAi, each with biological and technical replicates. Considering the variation between samples across experiments, data transformation of CI was performed and used for the CIGR. Normalization of the averaged CIGR over the time was performed based on a robust version of Z-score normalization. A t-test was used to determine the hits of genes playing an essential role in cell division with p-values ≤ 0.05 for the Z-score (Table 3.3).

To determine the gene function as activators or regulators in cell division, samples were compared with the experimental controls. The cell proliferation was enhanced by gene silencing suggesting that these genes potentially inhibit cell division i.e. regulator. In contrast, cells with arrested proliferation induced by gene silencing suggested that these genes were involved in dividing cells i.e. activator. For example, the effect of esiRNA ATL2 on the trace of CIGR over time was scored by -3; this means that the score is three times standard deviation below the score of population mean, suggesting a strong effect induced by gene silencing (Table 3.3). Analysis on the knock down of genes hypothesised to affect cell proliferation, 16 classified genes as activators that may play important roles in cell proliferation (Table 3.3). In addition, the knocking down of 3 unknown genes was found to regulate the cell proliferation (Table 3.3). For the known genes knocked down by RNAi screening, 7 and 4 genes were found to be involved in cell proliferation or regulated by cell proliferation respectively. Taken together, 19 of 42 (45.2%) tested unknown cell cycle-associated genes were found to be potentially involved in cell proliferation; 11 of 28 (39.2%) known cell cycle-associated genes were found important in dividing cells after RNAi screening.

	Activator			Regulator		
	Gene	**Z-score**	**p-value**	**Gene**	**Z-score**	**p-value**
Unknown	*ATL2*	-3	p < 0.001	*THAP10*	0.7	p < 0.01
	C17orf53	-3	p < 0.001	*ARHGAP11B*	0.8	p < 0.001
	ARHGAP11A	-2.5	p < 0.01	*GPN3*	0.8	p < 0.01
	FIGNL1	-2.5	p < 0.001			(3 genes)
	FAM111A	-2.4	p < 0.001			
	RIBC2	-2.3	p < 0.001			
	CCDC77	-1.7	p < 0.001			
	ZNF367	-1.5	p < 0.001			
	KIAA1549L	-1.4	p < 0.001			
	RPL39L	-1.3	p < 0.01			
	REEP4	-1.2	p < 0.001			
	GSTCD	-1.1	p < 0.001			
	FAM72B	-0.9	p < 0.001			
	TMEM194A	-0.9	p < 0.001			
	UBR7	-0.9	p < 0.001			
	DEPDC1B	-0.5	p < 0.001			
			(16 genes)			
Known	*CCNE1*	-1.9	p < 0.001	*AURKB*	0.8	p < 0.01
	ESPL1	-1.3	p < 0.001	*E2F7*	0.9	p < 0.01
	CEP85	-0.8	p < 0.01	*SPACA1*	1.1	p < 0.001
	CENPA	-0.6	p < 0.01	*ORC1*	1.3	p < 0.001
	KIF11	-0.5	p < 0.01			(4 genes)
	MAD2L1	-0.3	p < 0.05			
	GINS1	-0.2	p < 0.05			
			(7 genes)			

Table 3.3: Gene with unknown function and known cell cycle-associated genes found to be important in function and regulation of dividing cells. Averaged CIGR from transformed CI for each sample was then normalized, minimizing the diversity between whole dataset of RNAi screen. Score cut-off was determined by t test with p-value below 0.05. Genes with normalized Z-score below 0 were classed as being essentially associated with cell proliferation (i.e. activators). Genes with score above 0 were categorized as regulators.

Gene	Description	Human Protein Atlas	Function
RIBC2	RIB43A domain with coiled-coils 2	?	?
FAM72B	Family with wequence wimilarity 72, member B	?	?
ATL2	Atlastin GTPase 2	?	Golgi apparatus and ER morphogenesis
RPL39L	Ribosomal protein L39-like	?	Upregulated in ESCs and HCC tumours
ARHGAP11B	Rho GTPase activating protein 11B	?	Neocortex expansion
C17orf53	Chromosome 17 open reading frame 53	Cytoplasm	?
REEP4	Receptor accessory protein 4	Cytoplasm	ER clearance from metaphase chromatin
DEPDC1B	DEP domain containing 1B	Golgi apparatus	De-adhesion mitotic HeLa cells
CCDC77	Coiled-coil domain containing 77	Nuclear membrane	?
FIGNL1	Fidgetin-like 1	Nucleoli, Cytoskeleton	Homologous DNA repair
GPN3	GPN-loop GTPase 3	Nucleus, Cytoplasm	Required for proper localisation of RNA
GSTCD	Glutathione S-transferase, C-terminal domain containing	Nucleus	Expression in the human lung
UBR7	Ubiquitin protein ligase E3 component n-recognin 7 (putative)	Nucleus	Involve the spermiogenesis
TMEM194A	Transmembrane protein 194A	Nucleus but not nucleoli	?
ZNF367	Zinc finger protein 367	Nucleus but not nucleoli	Inhibit the cancer progression
THAP10	THAP domain containing 10	Nucleus but not nucleoli;Cytoplasm	THAP10 expression is regulated by promoter
KIAA1549L	KIAA1549-like	Nucleus but not nucleoli;Vesicles	?
FAM111A	Family with sequence similarity 111, member A	Nucleus, Nucleoli	Kenny-Caffey syndrome type 2
ARHGAP11A	Rho GTPase activating protein 11A	Nucleus, Nucleoli, Cytoplasm	Co-ordinate with p53 to regulate cell

Table 3.4: Description of genes identified by RNAi screening, subcellular location, and associated biological function according to public resources and published literatures. 19 unknown genes with hits were manually curated according to the subcellular protein localisation dataset of 'The Human Protein Atlas' project and literature mining [288]. The question mark (?) indicates that information was not available to achieve.

Cell proliferation affected by RNAi

In order to systematically demonstrate the effect of RNAi on cell proliferation, RTCA profiles were summarized into a number of groups. The grouping was based on the effect of RNAi on the CIGR classified by Z-score cut-off, which was categorized according to hits. In addition, according to the negative control releasing time point, the CIGR trace demonstrated where the maximal effect of RNAi affect was (Figure 3.11 and Table 3.3). To systematically summarize the RTCA profile and efficiently determine the effect of RNAi on cell proliferation, the trace of CI, CIGR, and CIGR with respect to the negative control over time were demonstrated side by side in Figure 3.11 across each RTCA profile.

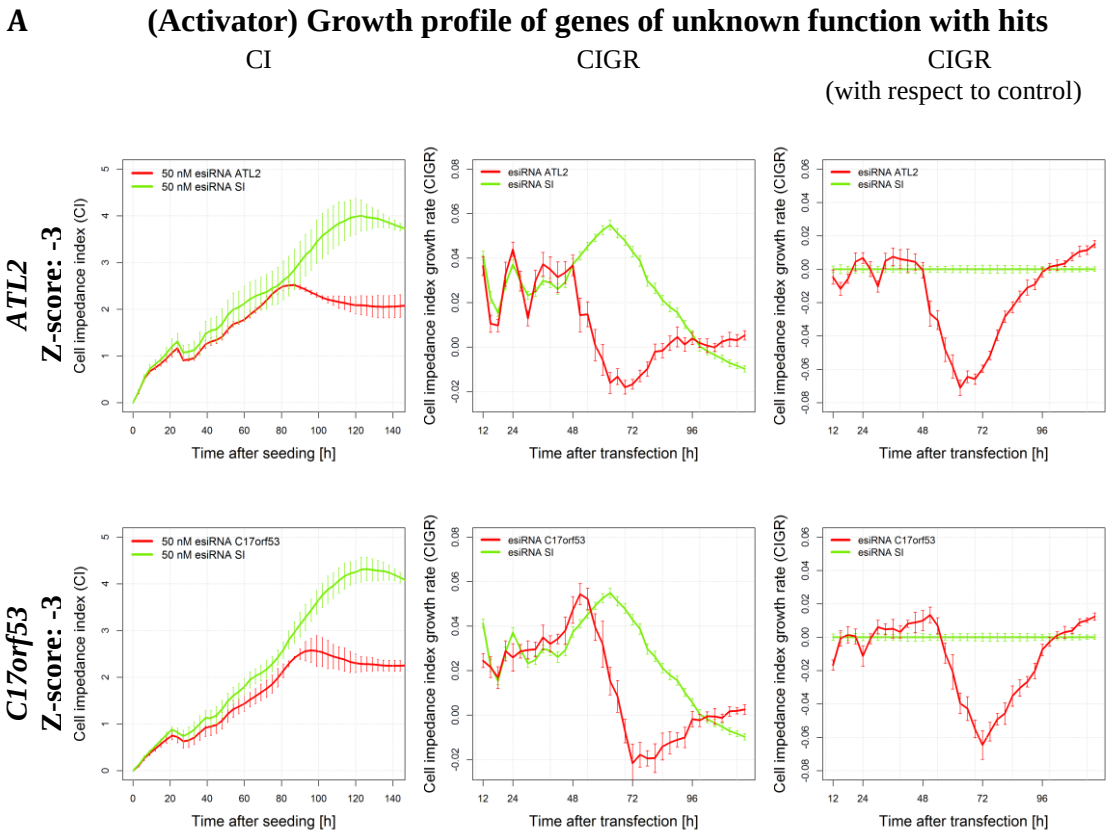
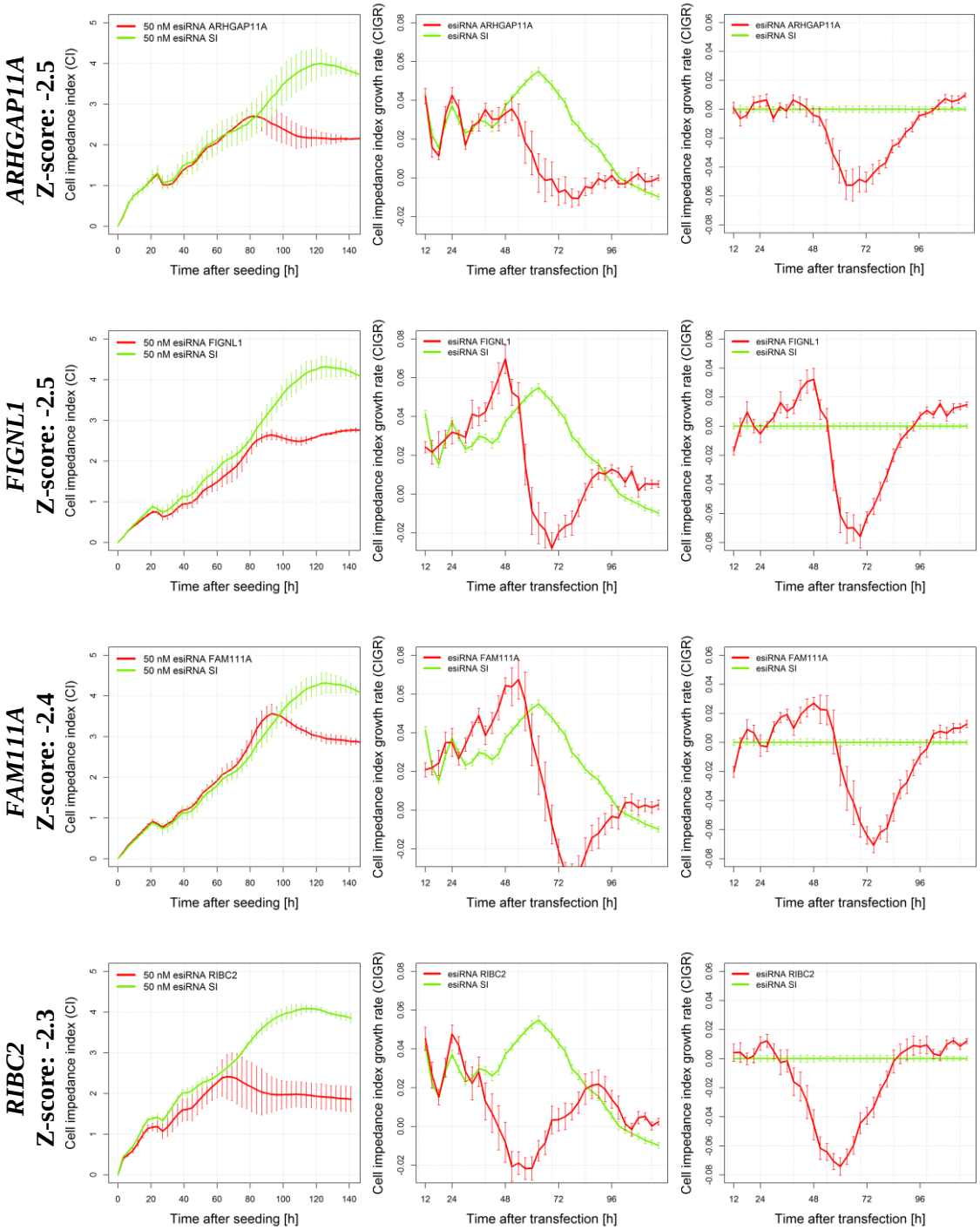


Figure 3.11: RTCA profile of the trace of CI, CIGR, and CIGR with respect to time. The trace of CI for each gene in RNAi screening is composed of technical triplicates in each 96 well E-plate and each E-plate was repeated three times. Plots of RTCA profiles were demonstrated in CI, CIGR, and CIGR with respect to negative control over time. For the column of CI over time, the transfection is at 24 h after cell plating. Red line indirectly represents that cell proliferation before and after transfection. Green line represents the negative control. Error bars represent mean ± SD. For the trace of CIGR and CIGR with respect to control over time, red line represents the CIGR of cells after transfection. Green line represents the negative control. Error bars represent mean ± SE. RTCA profile of tested genes were categorized based on the Z-score cut-off in **A)** growth profile of genes of unknown function as activators in cell proliferation, **B)** known cell cycle-associated genes as activators in cell proliferation, **C)** Genes with unknown function potentially functioning as regulators in cell proliferation, and **D)** known cell cycle-associated genes potentially functioning as regulators in cell proliferation.

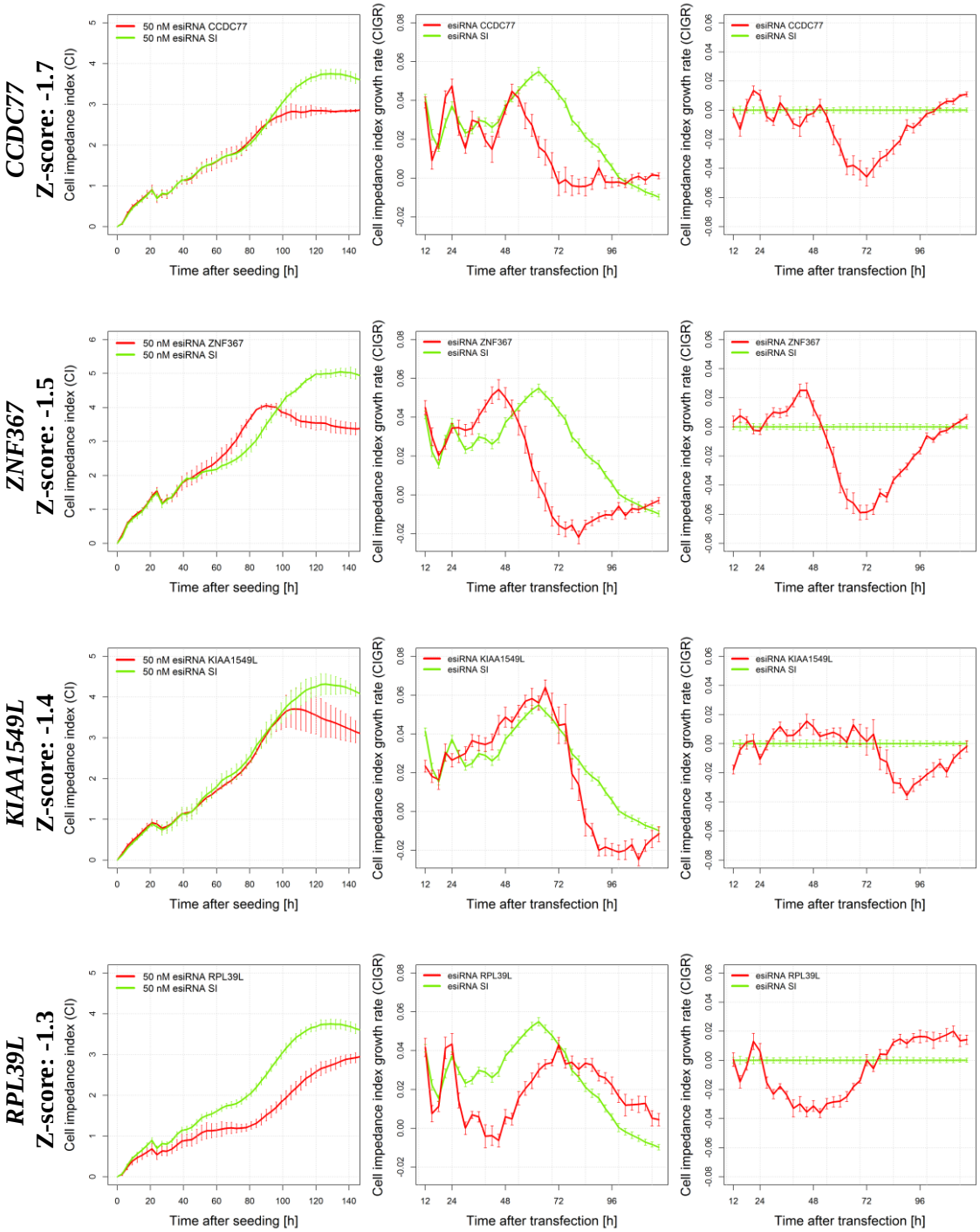
A **(Activator) Growth profile of genes of unknown function with hits**

CI CIGR CIGR (with respect to control)



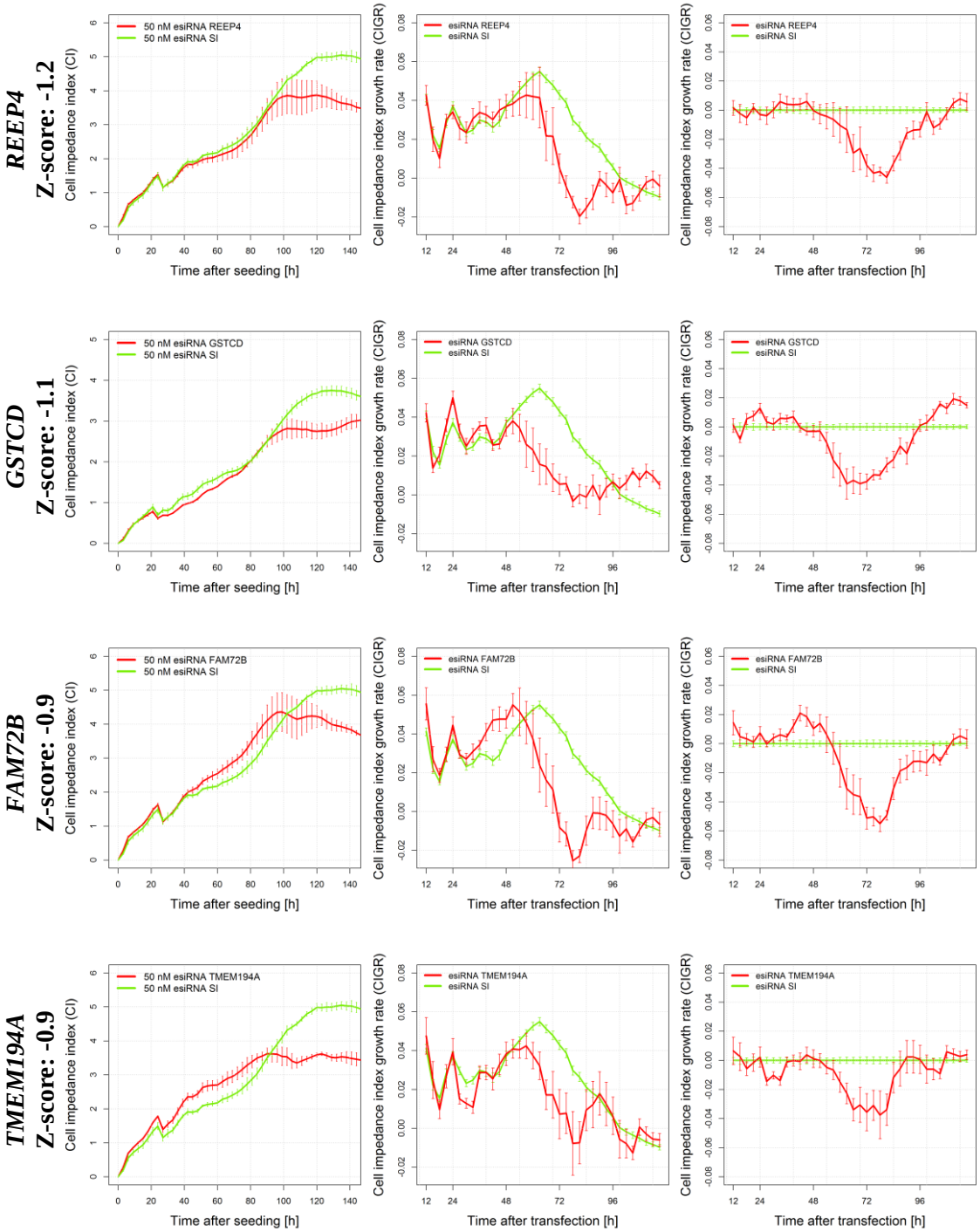
A **(Activator) Growth profile of genes of unknown function with hits**

CI CIGR CIGR (with respect to control)



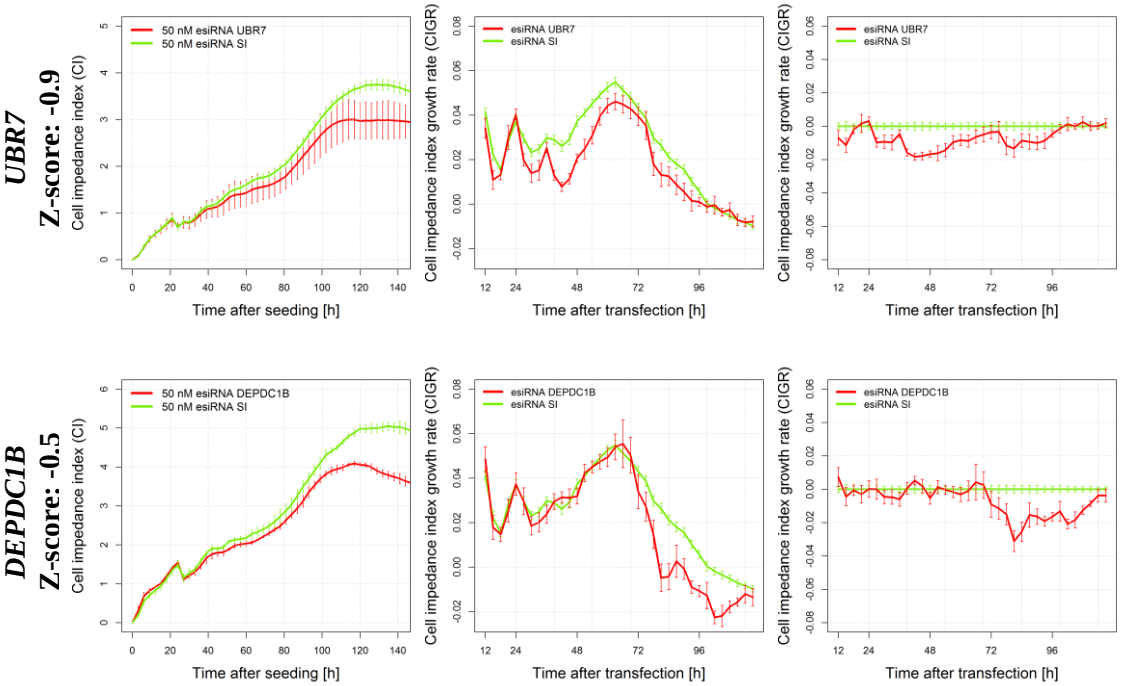
A **(Activator) Growth profile of genes of unknown function with hits**

CI CIGR CIGR (with respect to control)



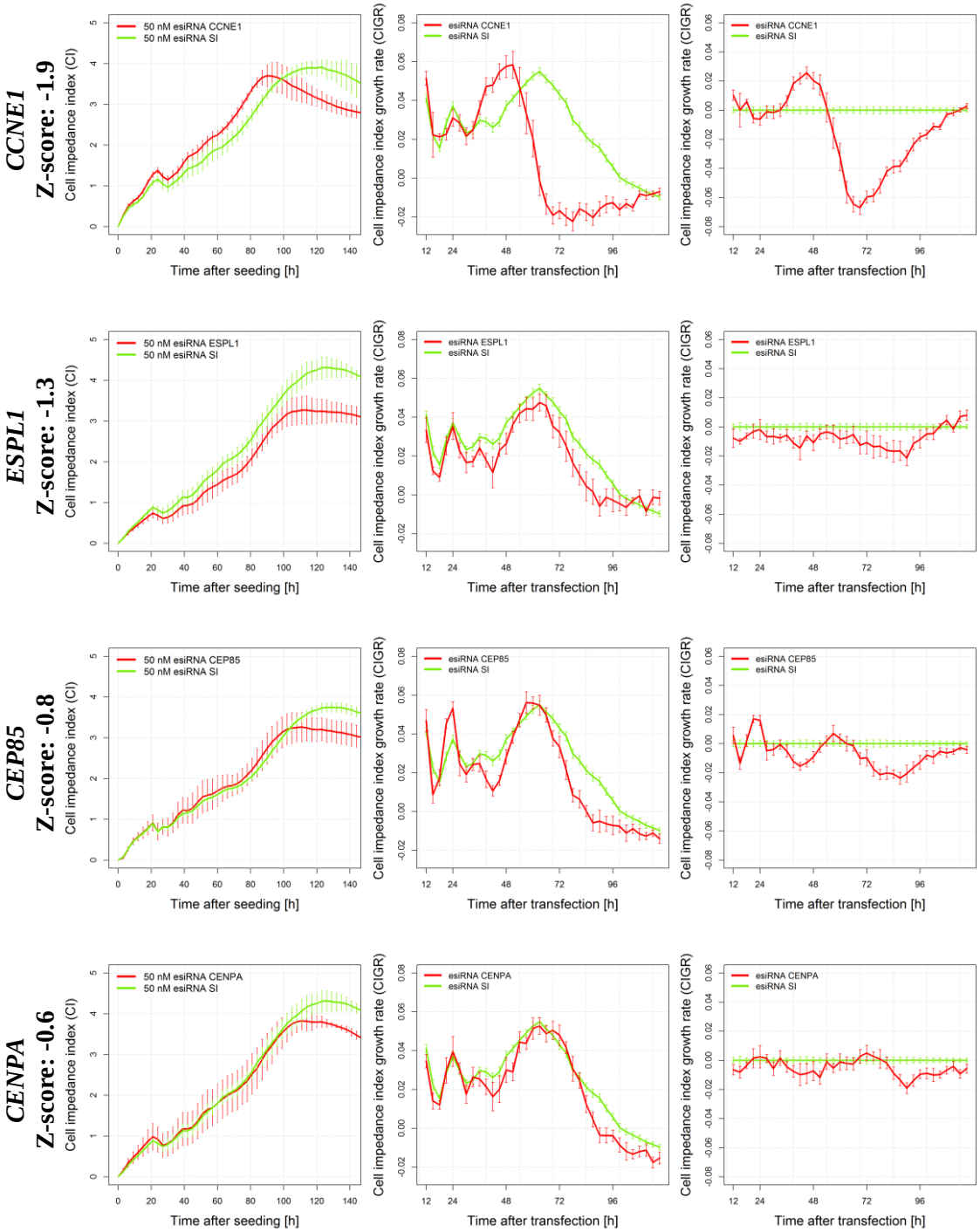
A **(Activator) Growth profile of genes of unknown function with hits**

CI CIGR CIGR (with respect to control)

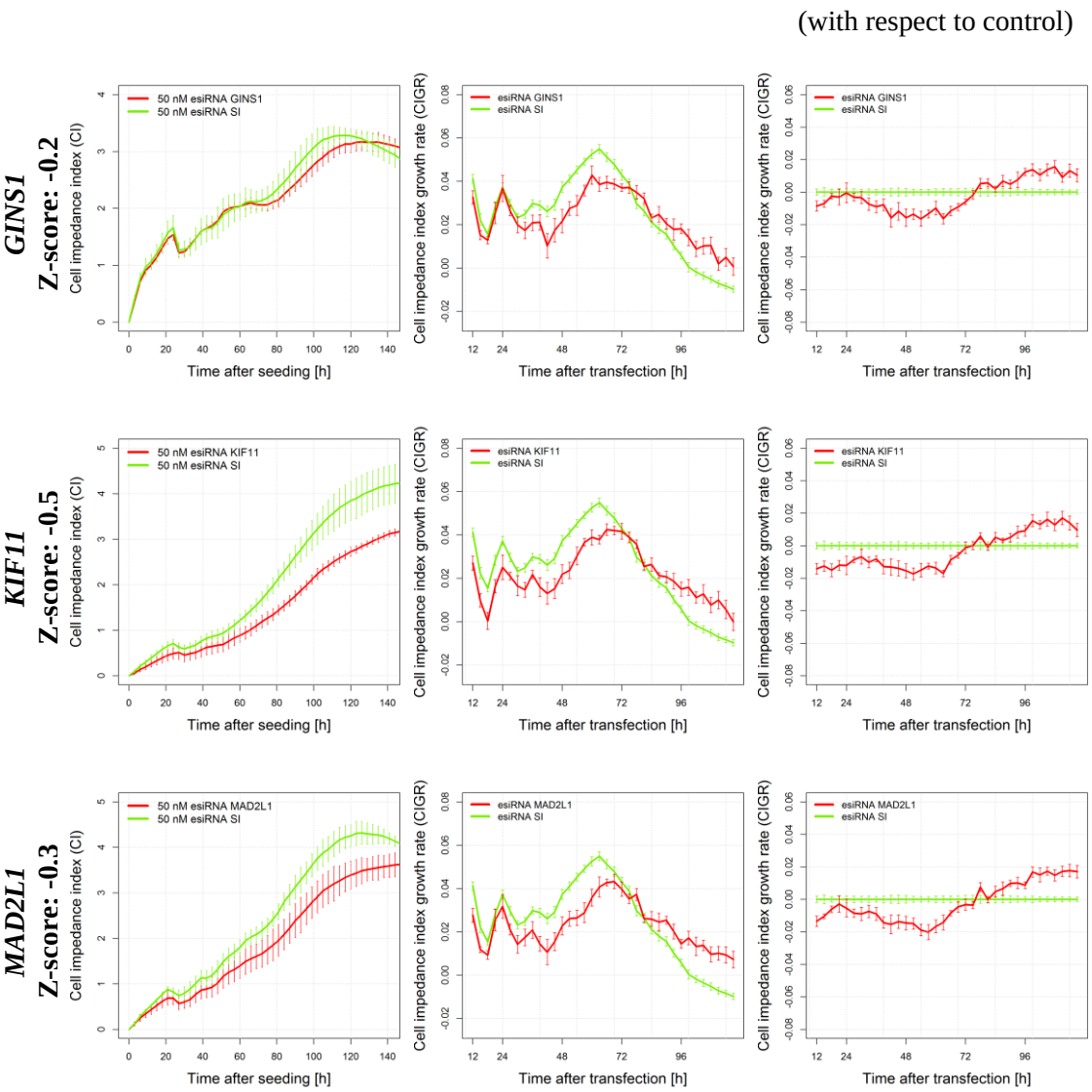


B **Known cell cycle-associated genes with hits**

CI CIGR CIGR (with respect to control)

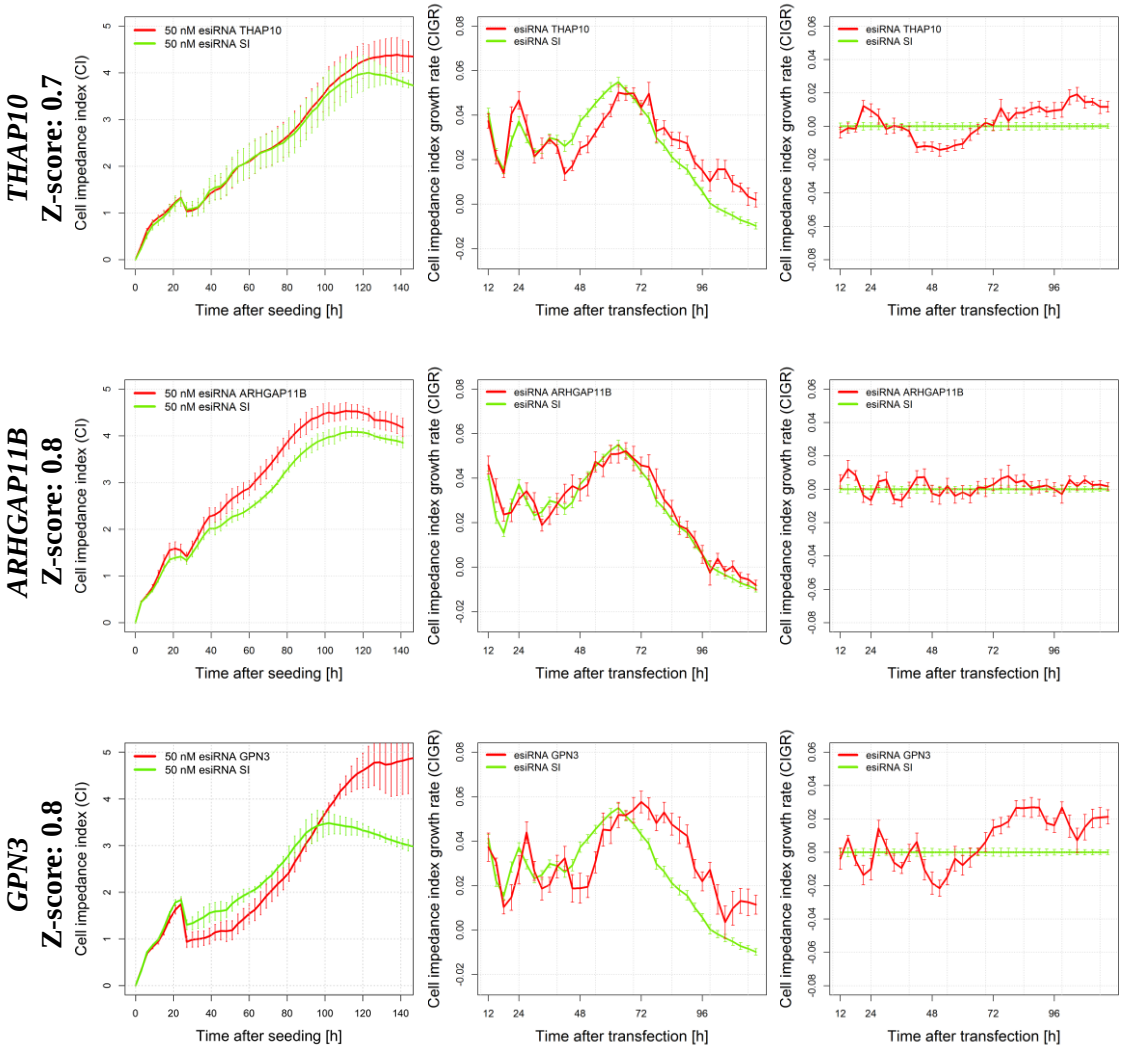


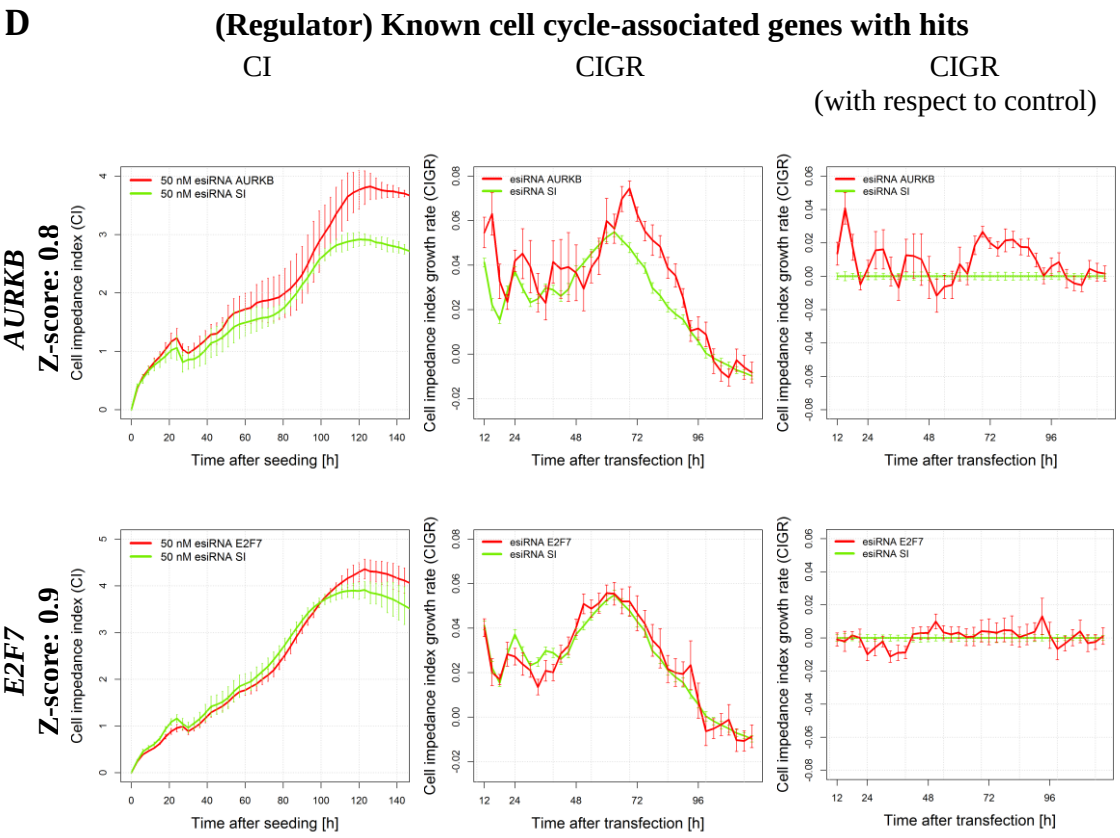
B **Known cell cycle-associated genes with hits**



C **(Regulator) Genes of unknown function with hits**

CI CIGR CIGR (with respect to control)





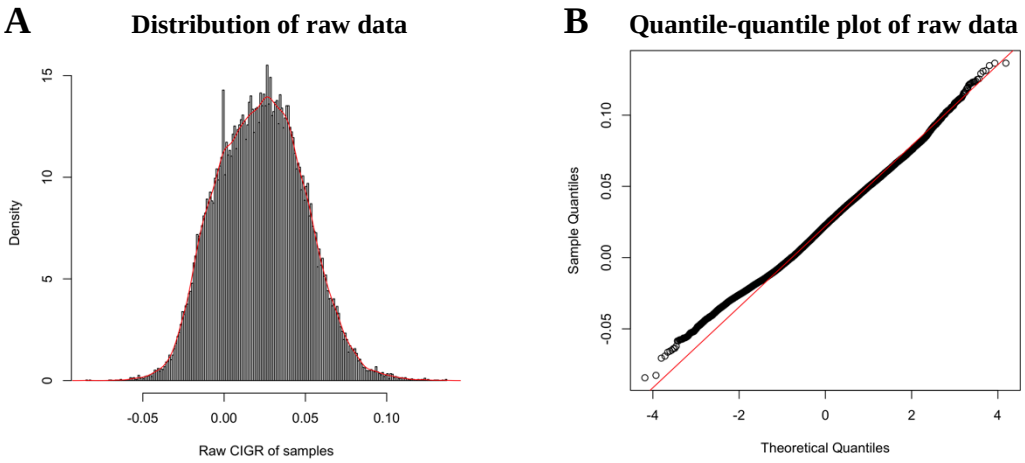


Figure 3.12: Distribution of CIGR in real time RNAi screening determined by histograms and quantile-quantile plots. A) Histograms with density plots represent the distribution of samples of biological triplicates of raw data of CIGR. **B**) The linearity of points between the quality of collected samples and theoretical distribution determines that there is no need to normalize collected data. The quantile-quantile is generally used to compare the normal distribution of collected samples against theoretical model, providing an assessment of "goodness of fit". If the plot is roughly linear, then the data are approximately normally distributed. The linearity of raw data distribution suggests that there was no need for normalization.

Robust knock down efficiency of gene silencing induced by esiRNA

RT-qPCR was used to evaluate whether cell proliferation arrest was induced by the targeted gene in the RNAi screen at each time point by analysing the genes with the greatest knock down efficiency, which were identified using Z-score cut-off. In comparison to the relative expression level of the housekeeping gene *GAPDH*, the relative expression levels of five genes were reduced by greater than 80%, including *REEP4* (83.4%), *CCDC77* (84.1), *FIGL1* (83.1%), and *ZNF367* (82.2%) (Figure 3.11).

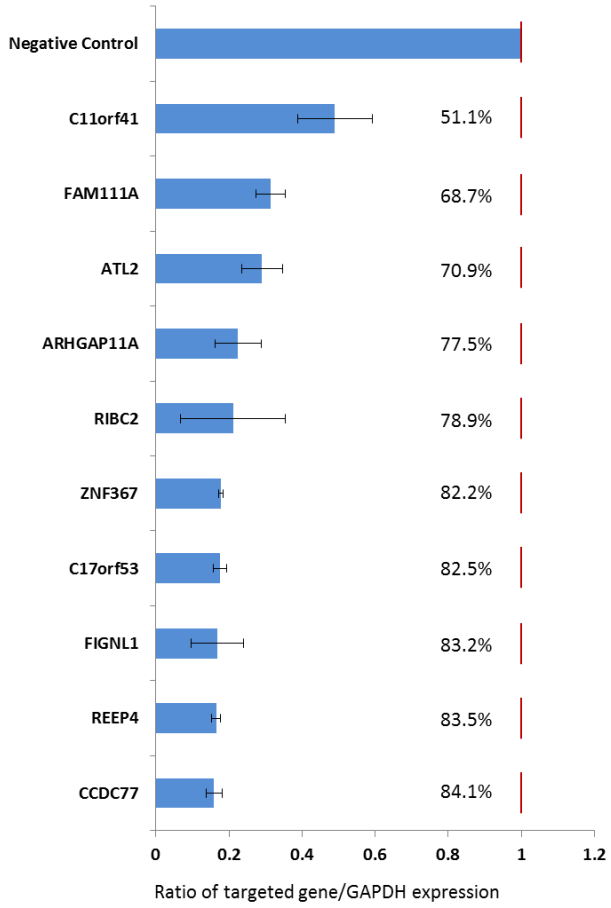


Figure 3.13: Knock down efficiency of genes identified as activators in in dividing cells. % of knock down efficiency was calculated between the expression level of targeted genes and housekeeping gene *GAPDH*. Data shown as mean ± SD of replicates (n = 3). Expression ratio was normalized relative to the housekeeping control (negative control), GAPDH.

There were 8 genes which had greater than 70% reduction in relative expression levels after transfection e.g. *CCDC77* (84.1%), *REEP4* (83.5%), and *FIGNL1* (83.2%) (Figure 3.13). There were only two genes with a knock down efficiency below 70%; *FAM111A* (68.7%) and *KIAA1549L* (51.1%) (Figure 3.13). This suggested that gene silencing provided robust knock down efficiency of targeted genes.

Discussion

To further investigate the role of a number of unknown cell cycle-associated genes, 42 unknown cell genes were screened by RNAi to see if knocking down their expression had any effect on cell proliferation.

Given that the knock down efficiency of RNAi was on average around 70% at 48 h following the transfection (Figure 3.8), a minimal period of real time RNAi monitoring has to be at least 72 h or longer to determine the effect on cell proliferation (Figure 3.10). After cell seeding, 24 h is the recommended time required before it is ready for transfection according to manufacturers' recommendation. Therefore, 96 h is the minimal time period required to provide sufficient CI measurements over time to determine the effect of RNAi on cell proliferation.

In addition, a sufficient seeding density for cell proliferation to arrive at the stationary phase between 96 and 120 h is also required. The length of exponential phase of CI over time is positively correlated with serially diluted seeding densities from 250 to 6,000 cells/cm^2 (Figure 3.5). In comparison with other seeding densities, only 6,000 cells/cm^2 density can provide a sufficient number of cells to arrive at the stationary phase in 72 h (96 h after cell plating) i.e. 100% cell confluency.

It is also important to note that a dramatically increased measurement of CI at the time point immediately following cell plating is induced by cells adhering to plate (Figure 3.5 red box). This was also observed by Zhang *et al.*, 2011 [115].

To perform the RNAi screening in the RTCA system, it is important to provide a culturing environment in which the cells can adhere to a 96 well E-plate, for example, by using a laminin coating. Laminin contributes to the structure of an extracellular matrix that enhances the cell adhesion. However, since the electrode on 96 well E-plate is a highly sensitive sensor, coating may also influence the accuracy of real time monitoring. The CI of cells plated on laminin coated and coating free plates were compared. In contrast to cells without coating, the trace of CI with laminin coating was reduced after 40 h following the initial cell plating (Figure 3.6). This suggests that the cells can adhere and proliferate without a coating of laminin on the E-plates.

In Figure 3.7A, the trace of CI over time showed that transfection reagent Lipofetamine® 2000 arrested the cell proliferation until the point where it was replaced by complete

medium. Zhang, *et al.*, 2011 also found that using liposome to conduct transfection may potentially result in cell proliferation arrest, resulting in synchronization of the cell cycle [115]. Both cationic liposome and magnetic nanoparticles can effectively conduct the transfection (Figure 3.8). However, it has been suggested that cationic liposome has a dose-dependent cytotoxicity on cultured human cells, such as macrophages [289-291]. Cationic liposome inhibits 3H-thymididine during DNA duplication, inhibiting cell proliferation[289].

In contrast, the trace of CI using silenceMag™ for transfection (Figure 3.7B) shows only a slight effect on cell proliferation after reagent replacement. The silenceMag™ was therefore considered the optimal transfection reagent for RNAi screening.

A comparison of knock-down efficiency using two methods of RNAi (siRNA vs. esiRNA)

In order to find the best conditions for RNAi screening that can robustly affect the cell proliferation, different combinations of transfection reagents and RNAi constructs were compared. The expression level between targeted and housekeeping genes is compared showing that the transfection of siRNA conducted by lipofectamine® 2000 has a stronger effect than esiRNA transfected by silenceMag™ (Figure 3.8). However, it is interesting that in the transfection reaction using silenceMag™, the knock down efficiency of gene silencing induced by siRNA is statistically similar to esiRNA.

Genes with significant sequence conservation to target genes may result in an off-target effect of RNAi induced gene silencing, which may lead to a misinterpretation to the importance of the target gene [140]. In addition, it has been suggested that near-perfect matches between targeted mRNA sequences and the antisense strands of the siRNA can influence RNAi induced gene silencing.

If off-targeting is not mediated due to a perfect match between the antisense strand of siRNA and the mRNA, then overall identity makes little or no contribution to determining whether the expression of a particular gene will be affected by given siRNA [143]. In this case off-targeting is mediated by perfect matches between the 2-7 or 2-8 bps (the seed region) and the 3’ UTR (untranslated region) of the mRNA [143]. Reducing the concentration of siRNA or timing of the transfection dose not completely minimise the off-target effects [140].

esiRNA induced gene knockdown shown in Figure 3.12, demonstrates a robust knock down efficiency around 70 to 80% in most of the cases where the knocking down of these genes

affected the cell proliferation. The esiRNA that can knock down targeted genes with a minimal off-targeting effect and was therefore used in RNAi screening [121].

Since the effect of RNAi on cell proliferation is determined by the trace of CI over time, it is important to minimize any potential side effects, such as cytotoxicity that will arrest dividing cells synchronising the cell cycle. In addition, cell viability assessment using flow cytometry demonstrated a high cell viability of cells transfected via silenceMag^{TM} (Figure 3.9). Cell viability analysis shows that the knocking down known cell cycle regulation *CCNB1*, *FOXM1*, and the non-essential gene *COL1A2* does not induce the cell death, when transfected using silenceMag^{TM} (Figure 3.9). The magnetofection was therefore selected to transfect cells in RNAi screening.

Ratio transformation of CI over time stabilizes the standard deviation between wells and plates to improve the comparison of RNAi across the RTCA profile

Given the optimized conditions for RNAi screening, CI profiles that changed over time were extracted from the RTCA system and normalized according to the manufacturers' instruction. Periodical maximum effect of RNAi on CI over times was used to determine the effect of knockdown genes on cell proliferation. The CI between plates cannot be compared directly as the variance of cell seeding, transfection, minor oscillation of CO_2 concentration, and temperature can all influence the trace of CI [115]. In order to compare the effect of RNAi across the RTCA dataset, normalization was required to minimize the diversity between wells and plates.

Following the manufactures' instructions, the trace of CI at the fixed time point of normalization was readjusted from the recorded value to 1. It was found that standard deviation increased between technical and biological triplicates as time progressed from the fixed time point. To overcome this, Zhang *et al.*, 2011 suggests using the cell index growth rate (CIGR) instead of the CI to determine the effect of RNAi [115]. The CIGR represents the rate of change of CI over time, indirectly reflecting the dynamic of cell proliferation affected by RNAi (Figure 3.11). In addition, the standard deviation of the CIGR did not accumulate following a fixed time point selected for normalization but rather stayed constant (Figure 3.11).

The trace of CIGR in comparison with a negative control determines the time window of RNAi effect on cell proliferation

Using CIGR over time to determine the effect of RNAi on cell proliferation, stabilized the diversity between wells and plates, allowing categorization of the RTCA profile based on the time window of maximum effect of RNAi. However, owing to the complexity of determining the difference between traces of CIGR that oscillate over time against controls, it is still difficult to accurately define the time window of maximum RNAi effect. Therefore, the profile of CIGR affected by RNAi is demonstrated with the respect to negative control.

It is interesting that in Figure 3.11, the trace of CIGR plotted with respect to the control showed two major populations of samples with a maximum effect of RNAi at 48 and 72 h after transfection. For example, the knock down of gene RPL39L, KIAA0586, ANKRD32, and ANP32E affected the value of CIGR over time at 48 h. On the other hand, the knock down of gene *ATL2*, *C17orf53*, *AHGAP11A*, *FIGNL1*, *FAM111A*, *ZNF367*, *CCDC77*, *GSTCD*, *FAM72B*, *REEP4*, and *TMEM194A* affected the CIGR over time at 72 h after transfection (Figure 3.11).

Half-life of protein influences the effect of RNAi

Knocking down of some known cell cycle associated genes that do not or only weakly affect cell proliferation. It is believed that this is due to the longer half-life of encoded proteins, which minimize the effect of gene silencing. In addition, by examining the high-throughput half-life data demonstrated by Yen *et al.*, 2008 [292], supportive evidence suggests that these are mostly identified as long life proteins, such as gene *E2F8*, *MCM10*, *PCNA*, and *CDC20*.

In contrast to the list of genes that were verified by RNAi using RTCA system, it is interesting that the knock down of some of the known cell cycle-associated genes didn't have an effect on cell proliferation e.g. *CCNE2*, the regulatory cyclin for CDK2, a known regulator that progresses the S-phase. However, it has been suggested that cell cycle can be progressed in mammalian cells without cyclin E activity by targeting both genes encoding cyclins E1 and E2 [293,294]. This potentially explains why the knocking down of gene *CCNE2* did not affect the rate of cell proliferation.

Limitation of RNAi and scoring cut-off

In order to determine the effect of gene silencing on cell proliferation, Z-score cut-off, an end-point score cut-off is widely used in typical high-throughput RNAi screening. This method was used to determine the hits of targeted genes based on their CIGR, providing a global scale of hits that determined the effect of RNAi between wells and plates [130-132,147,287]. To level the effect of RNAi on cell proliferation, which was indirectly monitored by highly sensitive RTCA system, normalization was required to minimize the diversity between wells and plates. For scoring the effect of the RNAi screen, the R package 'cellHTS2' which is developed for end-point high-throughput RNAi screen, was used to normalize average CIGR over time between wells and plates (Figure 3.11 and Table 3.3) [287].

Due to the limitation that a Z-score can only determine which genes in the RNAi screening are hit at a fixed time point, Zhang, *et al.,* 2011 suggested calculating the average of CIGR over time as a single value. Therefore averaged CIGR can be used to analyse the effect of gene silencing on cell proliferation [115]. However, adopting this scoring system instead of a typical end-point analysis might also lead to misinterpretation of the results of the RNAi screen.

In addition, RNAi is purely a loss-of-function technique aimed to target the gene of interest. The quality of this technique is affected by a number of conditions i.e. transfection efficiency, poor chemical stability allowing for degradation by RNases leading to low half-life, difficult to silence the targeted genes with high transcriptional activity in cells, and poor understanding of the true endogenous function of dsRNA molecules in vivo [295].

In summary, a real time esiRNA screening was carried out using silenceMag^{TM}, a cytotoxicity free transfection reagent to knock down unknown cell cycle-associated genes without off-target effects. RTCA was used to profile the effect of RNAi on cell proliferation based on electrical impedance as the output. The optimized culture system for NHDF cells provides a robust environment for a gene silencing experiment, which consists of optimised seeding densities, and an optimised length of time for cell proliferation tracking. Transformation of CI over time as CIGR stabilizes the standard deviation, which can be used to compare the effects of RNAi between wells and plates, demonstrated with respect to negative control; the Z-score cut-off method that may potentially miss the unknown weak RNAi effect on cell proliferation.

In addition, conditions that include the statistical analyses and various influences may determine the accuracy of results and could possibly miss weak effects, only highlighting the RNAi with strong effects on cell proliferation can be identified. Therefore, the design and analysis determine the genes that play important role in the cell proliferation.

Manual curation of genes to determine their role and location

Knowing the location of unknown proteins is the first step to predicting their role and function. The functional activities of proteins are associated with their subcellular localisation and protein-protein interactions. To predict putative role of target gene in cell proliferation, public resources including 'The Human Protein Atlas' and other published literatures were utilised in this analysis (Table 3.4) [288].

It was reported that C17orf53 and REEP4 appeared to be localised to the cytoplasm. REEP4 was found to be involved in the regulation of endoplasmic reticulum (ER) clearance during [296]. Little is known, however, about the putative role of C17orf53 in the cell proliferation. GSTCD and UBR7 were both reported to be localised to the nucleus. GSTCD was found to be highly expressed in the human lung tissues [297]. UBR7 was reported to be involved in the spermiogenesis [298]. However, there is little information available for these genes that is relevant to cell proliferation. Protein ZNF367 and TMEM194A were both reported to be localised to the nucleus but not the nucleoli. The expression of ZNF367 was reported to regulate the development of cancer progression [299]. In contrast, little information is available for the role of TMEM194A and additional experimental are required to identify the role of these genes in cell division (Table 3.4). Proteins THAP10 and KIAA1549L are also reported to be localised to the cytoplasm and vesicles respectively. The expression of THAP10 has been reported to be regulated by promoter hypermethylation in breast cancer [300]. However, little is known about the role of KIAA1549L (Table 3.4).

Other proteins localised to the nucleus including, FAM111A and ARHGAP11A which were both reported to be localised to the nucleus and nucleoli. In addition, GPN3 and ARHGAP11A were both reported to be localised to the nucleus and cytoplasm. GPN3 was known to be involved in the regulation of RNA polymerase II during the transcription [301]. A functional mutation of FAM111A was found to be associated with a skeletal disorder [302]. Interestingly, it has been suggested that ARHGAP11A influences cell proliferation by involving the p53 pathway (Table 3.4) [303].

Five of the proteins have no further information suggesting their position in cells i.e. RIBC2, FAM72B, ATL2, RPL39L, and ARHGAP11B. ATL2 has been reported to be linked to the

Golgi apparatus pathway and ER morphogenesis [304]. RPL39L has been reported to be highly expressed in embryonic stem cells and also suggested to be strongly associated with the hepatocellular carcinoma tumours (HCC) [305]. ARHGAP11B was reported to be involved in the development of human neocortex [306]. In contrast, little information is available for protein RIBC2 and FAM72B. CCDC77 appears to be localised to nuclear membrane according 'The Human Protein Atlas', however, further experimental approaches are required to identify its role in cell proliferation (Table 3.4).

In contrast to the proteins with poor information relevant to their involvement in cell proliferation, DEPEC1B was reported to be localised to the Golgi apparatus involved in the regulation of de-adhesion of mitotic HeLa cells. The expression activity of DEPDC1B was found to be up-regulated during the G_2-phase and degraded during mitosis which is consistent with the expression periodicity observed in our array dataset [307]. Protein FIGNL1 was suggested to be localised to the nucleus and cytoplasm in association with the DNA repair proteins [308].

In summary, the knockdown of 19 of 42 unknown genes appears to have an effect on cell proliferation. Protein DEPDC1B, FIGNL1, and ARHGAP11A appear to be involved in cell cycle-associated events and online resources suggested their position in the cell. However, 5 of the 19 unknown genes have no information of their subcellular location. Therefore, to further identify their position in the cells and study their activities in cell proliferation, in Chapter 4, a subcellular protein localisation was performed. In addition, evidence of subcellular localisation is required for pathway formation, which will be discussed in Chapter 5.

Chapter 4.

Subcellular localisation of putative cell cycle proteins

Introduction

In Chapter 3, evidence was presented that 19 genes of unknown function may have a role in cell proliferation. Since the start of this project, evidence has been published by other research groups that confirms the role of 3 of these genes in cell cycle regulation. To further investigate the association of these genes with the cell cycle, the subcellular localisation of the proteins they encode was studied.

The subcellular localisation of a protein can be highly informative as to its biological function. To further our understanding of the uncharacterised genes and their putative roles in cell proliferation, the expression of GFP tagged full length clones (i.e. ORFs) in Gateway vectors was used to identify their subcellular localisation.

The studies used full length protein-coding cDNA clones from the human ORFeome project version 8.1, a public resource containing 12,230 clonal isolates (i.e. clonally-derived and sequence-verified ORFs cloned as a set of Gateway Entry clones) representing at least 11,149 human genes [170]. Gateway cloning is a universal cloning method dependent on site-specific recombination properties of the bacteriophage lambda [172]. This vector system provides a rapid and efficient way to deliver DNA sequences into multiple vector systems for protein expression [173].

Using a mammalian host for protein expression ensues the recombinant proteins undergo normal modifications at the post-transcriptional stage e.g. NHDF or HEK293T [309]. Of the cell lines used to express mammalian proteins, HEK293T are by far the most widely used cells [309]. This is because they stably synthesise SV40 T-antigen for effective plasmid expression. In comparison with NHDF, HEK293T has exhibited a transfection efficiency of between 95 and 100% [310]. Considering the time required to optimise the protein localisation experiment using NHDF, HEK293T were therefore chosen as an alternative platform for optimisation for protein subcellular localisation.

Protein localisation studies on 12 of the ‘hits’ i.e. genes of unknown function shown to cause a decrease in proliferation when knocked down by RNAi screening) from Chapter 3 (clones for the other 7 genes were not available) and another 8 genes identified by network structure

analysis were also included. Transfection of HEK293T cells was performed with expression clones containing the ORFs with GFP fused to either the N- or C-terminal.

Materials and methods

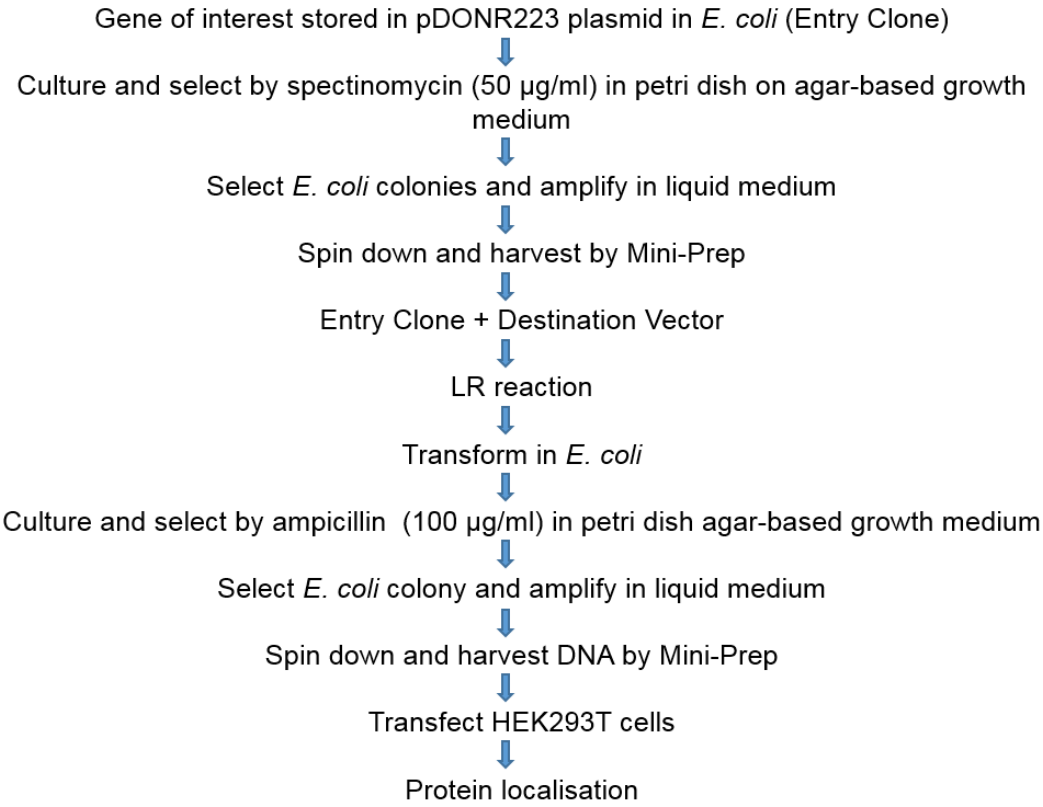


Figure 4.1: Flow chart of experimental steps in the protein localisation analyses.

Table 4.1 includes a list of genes of unknown function shown to reduce proliferation when knocked down by RNAi, as discussed in Chapter 3 and other genes of interest. Known cell cycle genes of interest were expressed to demonstrate their subcellular localisation for comparison with the results of the genes with unknown function.

	Gene	Description
Gene of unknown function identified by RNAi screening	*ATL2*	Atlastin GTPase 2
	C17orf53	Chromosome 17 open reading frame 53
	ARHGAP11A	Rho GTPase activating protein 11A
	FIGNL1	Fidgetin-like 1
	FAM111A	Family with sequence similarity 111, member A
	RIBC2	RIB43A domain with coiled-coils 2
	KIAA1549L	KIAA1549 like
	RPL39L	Ribosomal protein L39-like
	REEP4	Receptor accessory protein 4
	GSTCD	Glutathione S-transferase, C-terminal domain containing
	UBR7	Ubiquitin protein ligase E3 component n-recognin 7 (putative)
	DEPDC1B	DEP domain containing 1B
Genes of interest from the cell cycle list	*ANLN*	Anillin, actin binding protein
	DEPDC1	DEP domain containing 1
	FAM111B	Family with sequence similarity 111, member B
	KIAA0101	Proliferating cell nuclear antigen-associated factor
	KNSTRN	Kinetochore-Localised Astrin/SPAG5 Binding Protein
	KPNA2	Karyopherin alpha 2 (RAG cohort 1, importin alpha 1)
	RACGAP1P	Rac GTPase activating protein 1 pseudogene
	TCF19	Transcription factor 19
Known cell cycle gene	*CEP55*	Centrosomal Protein 55kDa
	CEP57	Centrosomal Protein 57kDa
	CEP72	Centrosomal Protein 72kDa
	CEP85	Centrosomal Protein 85kDa
	PCM1	Pericentriolar Material 1

Table 4.1: List of genes used for protein localisation by expressing GFP-fusion protein in HEK293T cells. This table includes genes of unknown function identified by RNAi screening, genes identified by network structure analysis only (i.e. genes on the cell cycle list), and 5 centrosome-associated genes. Transforming Acidic Coiled-Coil Containing Protein 3 (*TACC3*) tagged with GFP was used to determine the transfection efficiency in HEK293T cells.

Generation of entry clone by the Wiemann laboratory

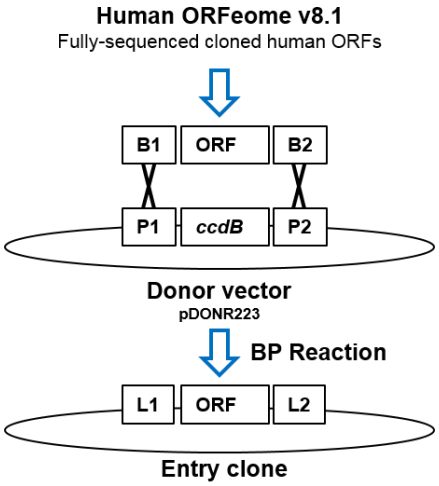


Figure 4.2: Fully-sequenced cloned human ORFs were flanked by combinational sites B1 and B2 and cloned into the P sites of pDONR223 using the BP reaction. This was performed by the Wiemann laboratory to produce the ORFeome version 8.1 library.

ORFs cloned in the pDONR223 plasmid were received as a gift from Prof. Stefan Wiemann, German Cancer Research Centre, Heidelberg (Figure 4.2) [170]. *E. coli* containing the plasmids were stored in 96-well plates. Inoculation from the 96 well plates were grown up in 15 ml tubes with 5 ml Luria Broth (LB) medium and incubated at 37ºC shaker overnight with aeration. Plasmids were harvested using the PureLink® Quick Plasmid Miniprep Kit (Life technologies, Paisley, UK) according to manufacturer's instructions and quantified on the ND-1000 Spectrohotometer (Nano-Drop Technologies, Wilmington, North Carolina, USA).

To prepare an expression clone, 150 ng of entry clone and 150 ng of destination vector (Gateway® pcDNA™-DEST53 or -DEST47 Vector, Life technologies) were mixed with 1 µl ddH_2O, and LR clonase II enzyme (Life technologies) and incubated at 25ºC for 1 h. 1 µl (2 µg/ µl) of proteinase K was then added to stop the activity of the enzyme (LR clonase II) by incubating for 10 min at 37ºC (Figure 4.3).

Sub-cloning the gene of interest from the entry clone (pDONR223) into a destination vector using the LR reaction to produce an expression clone

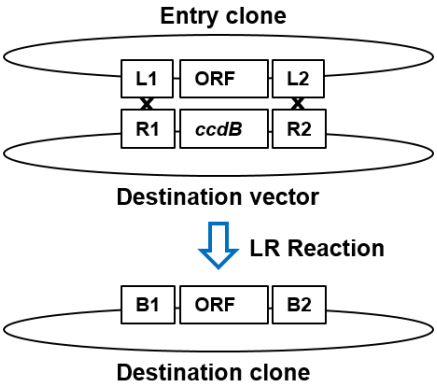


Figure 4.3: ORFs flanked by L1 and L2 sites were transferred from entry clones into *R1::ccdB::R2* destination vectors by using the LR reaction.

Transformation

For transformation, competent DH5α cells stored at –80ºC were thawed on ice. 1 µl of recombined plasmids were mixed with 50 µl competent cells. The DNA mixture was then kept on ice for 30 min. Each tube was given a heat shock by placing the tube in a 42ºC water bath for 45 seconds and then transferred to ice for 2 min. 950 µl LB media (without antibiotics) was added to each tube and incubated in a shaker for 1 h at 37ºC. Cells were then spun down at 3,000 rpm for 1 min, and the pellets re-suspended in 100 µl LB media. Re-suspended cells were then plated on to a 10 cm diameter LB agar plate with ampicillin (100µg/ml). Plates were incubated at 37ºC overnight. Single colonies were selected and grown up in a 15 ml tube with 5 ml LB containing ampicillin (100 µg/ml) for 17 h in a shacking incubator at 37ºC in preparation for the harvesting of expression clones.

Transfection of cells for protein localisation

For the transfection of HEK293T cells plated on coated microscope coverslips, the medium in the 24 well plates was replaced with the same volume (i.e. 500 µl) of antibiotic free medium. The transfection reagent K2® Transfection System (Biontex Laboratories, Munich, Germany) was used to deliver the expression clone. The transfection efficiency was calculated from repeats (n = 3) according to the number of HEK293T cells expressing TACC3 tagged with GFP. GFP expression was inspected by visual assessment at 1000x magnification.

Poly-L-Lysine coating and cell plating on cover slips

Poly-L-Lysine is commonly used to coat microscope slips to improve cell adherence [311]. Ultraclean 13 mm round glass coverslips were placed in 24-well plates, allowed to dry for 10 min, coated with 300 µl poly-L-Lysine at 100 µg/ml and incubated for 2 h at room temperature (Sigma-Aldrich, UK). HEK293T cells were then added to the 24 well plates with 500 µl DMEM (Dulbecco's Modified Eagles Medium, Sigma-Aldrich), including 10% (v/v) Fetal Bovine Serum (FBS) (GE Healthcare, Little Chalfont, UK), antibiotics (25 U/ml penicillin and 25 µg/ml streptomycin, Life technologies, UK), Non-Essential Amino Acids (NEAA) (1X) (Life technologies, UK), glutmax (1x) (Life technologies, UK), and fungizone antimycotic (1.25 µg/ml) (Life technologies, UK). The coated cover slips were growth with HEK293T cells were incubated in a 37^{o}C incubator overnight. The standard culturing conditions of HEK293T cells, are detailed in Chapter 2 materials and methods.

Labelling nuclei and cytoskeletal actin with DAPI and Phalloidin

For staining of HEK293T cells, the culture medium was removed and the cells were fixed in 4% paraformaldehyde (PFA) (Fisher Scientific, Leicestershire, UK) for 10 min. PFA was then replaced with PBS with 0.1% Triton X-100 (Sigma-Aldrich) and the sample stored at 4°C overnight. 200 µl of phalloidin (i.e. 7.5 units) was then added to replace the PBS (0.1% Triton X-100) for 30 min. Coverslips stained by phalloidin were then washed 3 times in PBS before adding 200 µl DAPI (300 nM) for 10 min. To mount on slides, the coverslips were washed 3 times in PBS. 5 µl Vectorshield (containing DAPI, Vector laboratories, Peterborough, UK) was placed on the frost-free slides before the coverslips were gently lowered onto each slide. 4 spots of nail polish were placed around the coverslips and allowed to dry for 10 min. The edges of the coverslips were then completely sealed with nail polish.

Microscopic inspection

Slides were inspected using a Nikon e-C1 confocal scanning laser microscope (Nikon Instruments, Surrey, UK). For each slide, 3 images were taken at 1000x magnification. Nikon EZ-C1 software (Nikon Instrument, Surry, UK) was used to control the 2-D optical sections to provide images from each channel with minimal spectral overlap. The stains and laser/filter combinations used were: phalloidin stain (excitation 543 nm, emission BandPass 605/75 nm), DAPI stain (excitation 408 nm, emission BandPass 461/30nm), and GFP (excitation 488 nm, emission BandPass 515/30 nm). For Z-stacks, optical sections were 1.6 µm.

Results

In Chapter 3, evidence was presented that 19 genes with unknown function play a role in cell proliferation. In order to further investigate their role in cell division, the subcellular protein localisation of 12 of the 19 proteins has been analysed by expressing them as GFP-fusion proteins. Known cell cycle-associated proteins and other proteins of interest involved in interphase or mitosis were also expressed fused to GFP to demonstrate their subcellular localisation and expression compared to published literature (Table 4.2B and C).

Table 4.2 summarises the subcellular localisation of proteins of interest expressed as fusions with GFP. Proteins were expressed with GFP located at their N- [using ORF cloned in pdDNA-DEST53 plasmid] or C-terminal [using ORF cloned in pcDNA-DEST47 plasmid]. Table 4.2A shows the 12 of 19 genes with unknown function expressed in HEK293T cells.

ATL2 protein tagged with GFP at the N-terminal was expressed in HEK293T cells. In interphase, the expression of ATL2 shows three spots proximal to the nucleus with weaker granular expression throughout (Figure 4.5A). In prometaphase, the localisation of ATL2 expression shows these discrete spots in the cytoplasm. These cells were representative of other cells expressing GFP in the samples (Figure 4.5B).

C17orf53 tagged with GFP was expressed in HEK293T cells. In interphase, the expression of C17orf53 with both N- or C-terminal GFP tagged protein shows a single spot proximal to the nucleus in the cytoplasm or throughout the cell (Figures 4.6A, B, C, and D). During prophase, the expression of C17orf53 tagged with the N-terminal shows three spots proximal to the nucleus (figure 4.6E). In prometaphase, the expression of C17orf53 tagged with GFP at the C-terminal shows a number of spots grouped in several clusters in the cytoplasm (Figure 4.6F). In telophase, the expression of C17orf53 tagged with GFP at the C terminal shows spots proximal to the chromosomes and weaker granular expression throughout the cell except in the condensed chromosomes (Figure 4.6G)

ARHGAP11A tagged with GFP at the N-terminal was expressed in HEK293T cells. In interphase, the expression of ARHGAP11A shows a granular expression throughout the nucleus but not the nucleoli (Figure 4.7A). During metaphase and cytokinesis, the granular expression of ARHGAP11A appears to be co-localised to the metaphase chromosomes with weaker granular expression throughout the cytoplasm (Figures 4.7B and C).

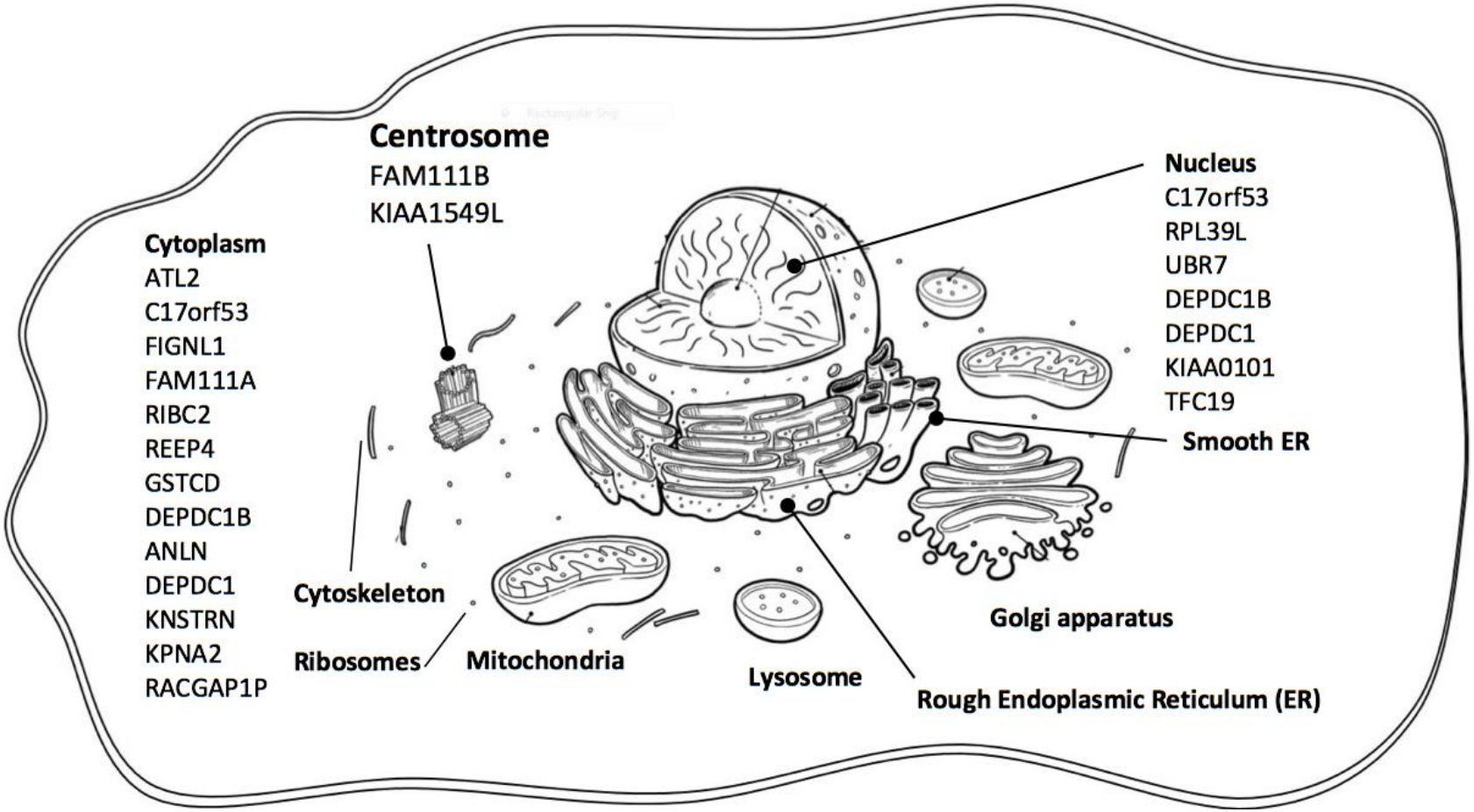


Figure 4.4: Summary of the subcellular protein localisation in an interphase cell. Protein FAM111B and KIAA1549L appear to be localised to the position in association with the centrosome during interphase. Diagrams were modified based on figures courtesy of http://www.timvandevall.com/wp-content/uploads/animal-cell-labeled.png.

A Gene of unknown function identified by RNAi screening

Gene	Agreement between N- and C-terminal expression	GFP	Interphase	Mitosis
ATL2	Yes	N-	Two or three spots with weaker granular expression in the cytoplasm	Prophase/ Two or three spots of expression in the cytoplasm
		C-	Not performed	
C17orf53		N-	Spots with weaker granular expression in the cytoplasm and nucleus	Prophase/ Two or three spots with weaker granular expression in the cytoplasm
		C-	Spots of expression in the cytoplasm and nucleus	Prometaphase, Telophase/ Spots with weaker granular expression in the cytoplasm
ARHGAP11A		N-	Granular expression in the nucleus but not the nucleoli	Metaphase, Cytokinesis/ Localised to the chromosome with weaker granular expression in cytoplasm
		C-	Not performed	
FIGNL1		N-	One or two spots with weaker granular expression in the cytoplasm	Prophase, Prometaphase/ One or two spots with weaker granular expression in the cytoplasm
		C-	Not performed	
FAM111A		N-	One or two spots with weaker granular expression in the cytoplasm	Prophase/ Two or three spots of distinct expression with weaker granular expression in the cytoplasm
		C-	A cluster of spots in the nucleus and cytoplasm	
RIBC2		N-	One or two spots in the cytoplasm	Prophase, Prometaphase, Telophase/ Spots with weaker granular expression in the cytoplasm
		C-	Not performed	
KIAA1549L		N-	One or two spots with weaker granular expression in the cytoplasm	
		C-	Granular expression in the nucleus but not in the nucleoli and the cytoplasm	
RPL39L		N-	Spots in the nucleus	
		C-	Not performed	
REEP4		N-	Throughout the cytoplasm	Two or three spots with weaker granular expression in the cytoplasm
		C-	Two or three spots with weaker granular expression in the cytoplasm	Prophase/ Numerous of spots of expression in the cytoplasm
GSTCD		N-	Throughout the cytoplasm but appears to be uneven expression with some spots	
		C-	Not performed	

Genes of interest from the cell cycle list

	Gene	Agreement between N- and C-terminal expression	GFP	Interphase	Mitosis
A	**UBR7**	Yes	**N-**	Nucleus but not in the nucleoli with granular expression	
			C-	Not performed	
	DEPDC1B		**N-**	Numerous spots with weaker granular expression in the nucleus and cytoplasm	
			C-	Granular expression in the cytoplasm and the nucleus	
B	**ANLN** (Cytokinesis)		**N-**	Throughout the cytoplasm but appears uneven on expression with some spots	
			C-	Not performed	
	DEPDC1 (Cytokinesis)		**N-**	Nucleoli but not in the nucleus/cytoplasm	
			C-	Not performed	
	FAM111B (Cell cycle regulation)		**N-**	One or two spots of distinct expression with weaker granular expression in cytoplasm	
			C-	One or two spots of distinct expression with weaker cytoplasm granular expression	
	KIAA0101 (DNA repair)		**N-**	A thick dashed lines in nucleus but not in the nucleolus	
			C-	Numerous dots in nucleus with weaker granular expression	
	KNSTRN (Cytokinesis)		**N-**	Throughout the cytoplasm but uneven with some spots	
			C-	Not performed	
	KPNA2 (Putative association)		**N-**	One or two spots of expression in the cytoplasm	
			C-	Not performed	
	RACGAP1P (Unknown function)		**N-**	Throughout the cytoplasm but appears uneven with some distinct spots and weaker cytoplasmic granular expression	Prometaphase/ Throughout the cytoplasm but appears uneven on expression with some distinct spots and weaker cytoplasmic granular expression
			C-	Not performed	
	TCF19 (Putative association)		**N-**	Numerous spots of distinct expression with weaker nuclear granular expression	
			C-	Not performed	

C

Known cell cycle gene

Gene	Agreement between N- and C-terminal expression	GFP	Interphase	Mitosis
CEP55		N-	Cytoplasm	Metaphase
		C-	Not performed	
CEP57		N-	Cytoplasm	Prometaphase & Metaphase
		C-	Not performed	
CEP72		N-	Cytoplasm	Telophase
		C-	Not performed	
CEP85		N-	Not performed	
		C-	Cytoplasm	Prophase & Prometaphase
PCM1		N-	Cytoplasm	
		C-	Not performed	

Table 4.2: Summary of the subcellular protein localisation. A) Genes with unknown function important to cell proliferation identified by RNAi screening. **B)** Genes of interest are thought to be involved in cell proliferation. **C)** Known genes expressed in association with the centrosome.

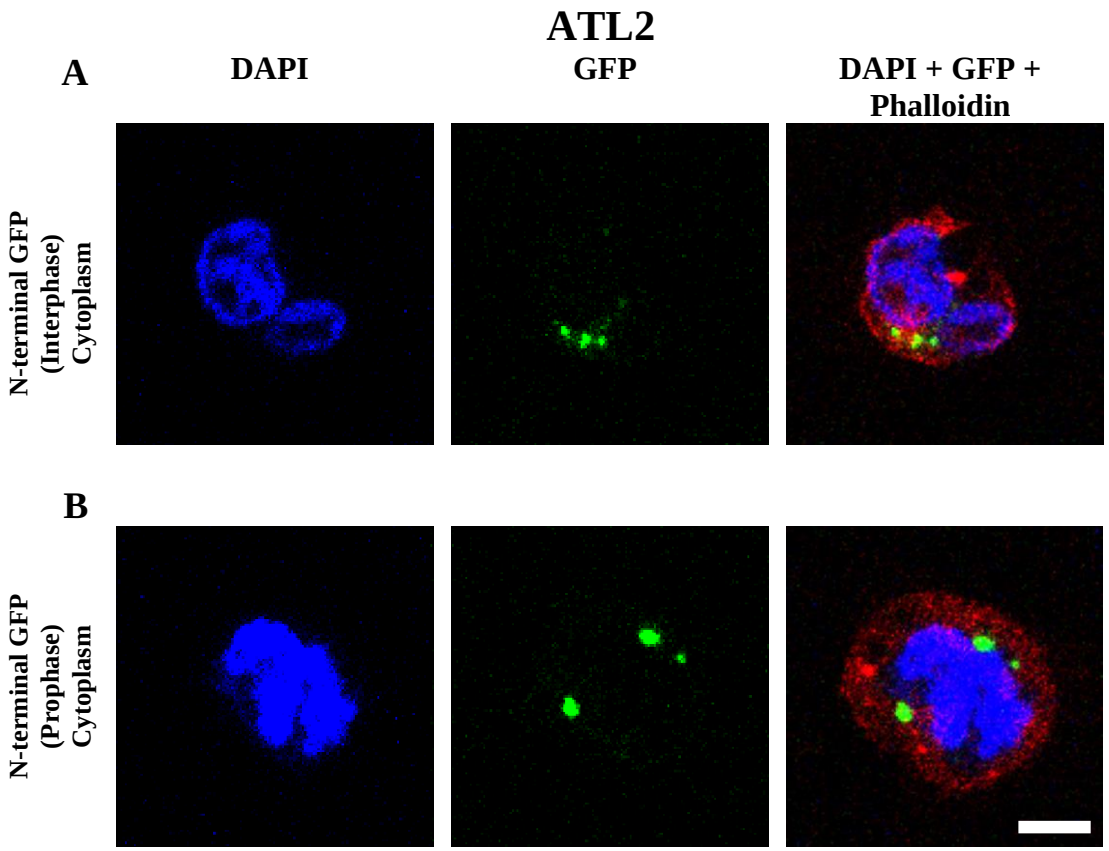
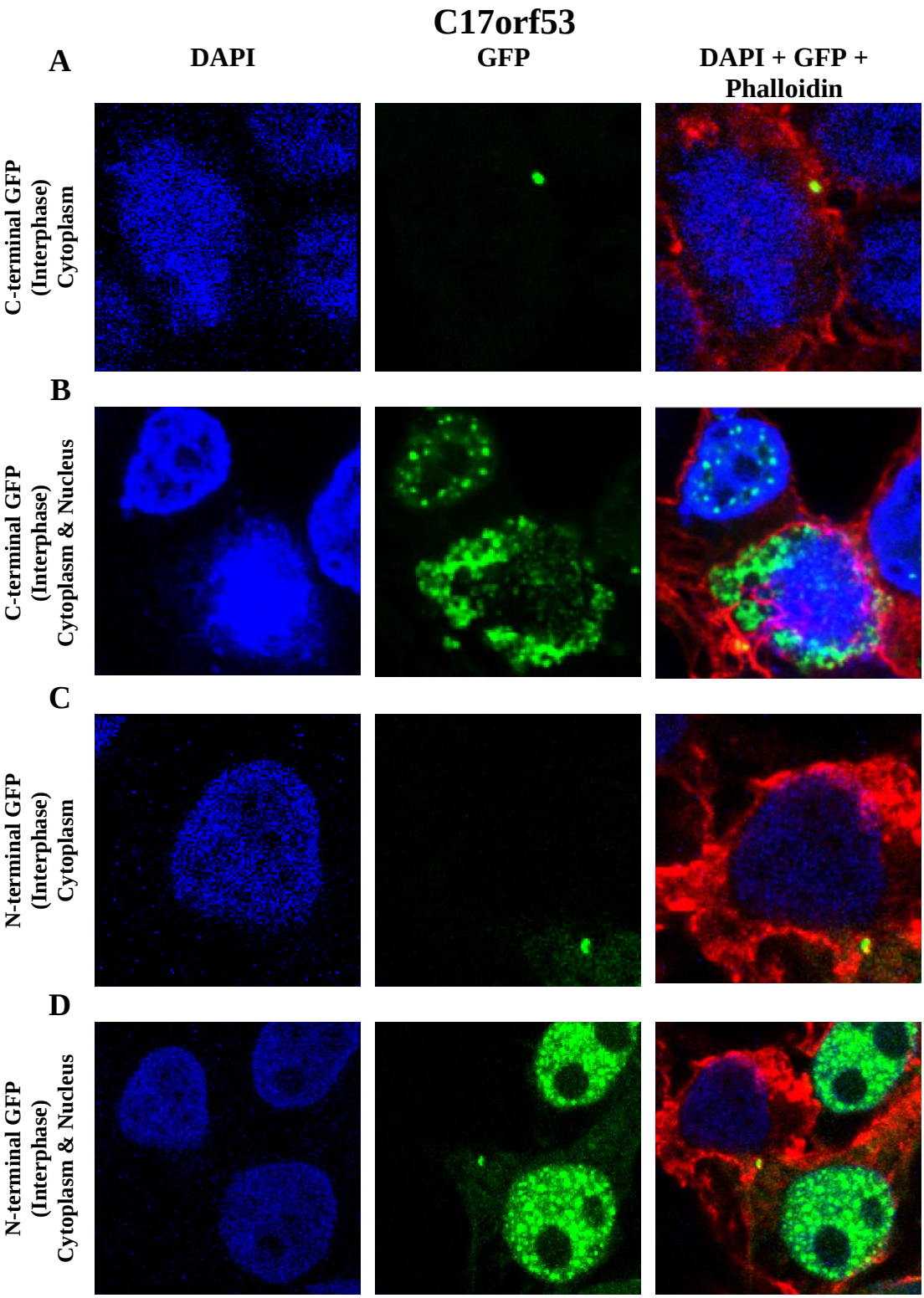


Figure 4.5: Subcellular localisation of ATL2 in HEK293T cells. The left column shows DAPI staining of the nucleus. The middle column shows HEK293T cells expressing ATL2 with N-terminal GFP tag. The right column shows the merged images of GFP expression, labelling with DAPI, and Phalloidin staining. **A)** Expression of ATL2 tagged with GFP at the N-terminal during interphase. **B)** Expression of ATL2 tagged with GFP at the N-terminal during prophase. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

Genes of unknown function identified by RNAi screening



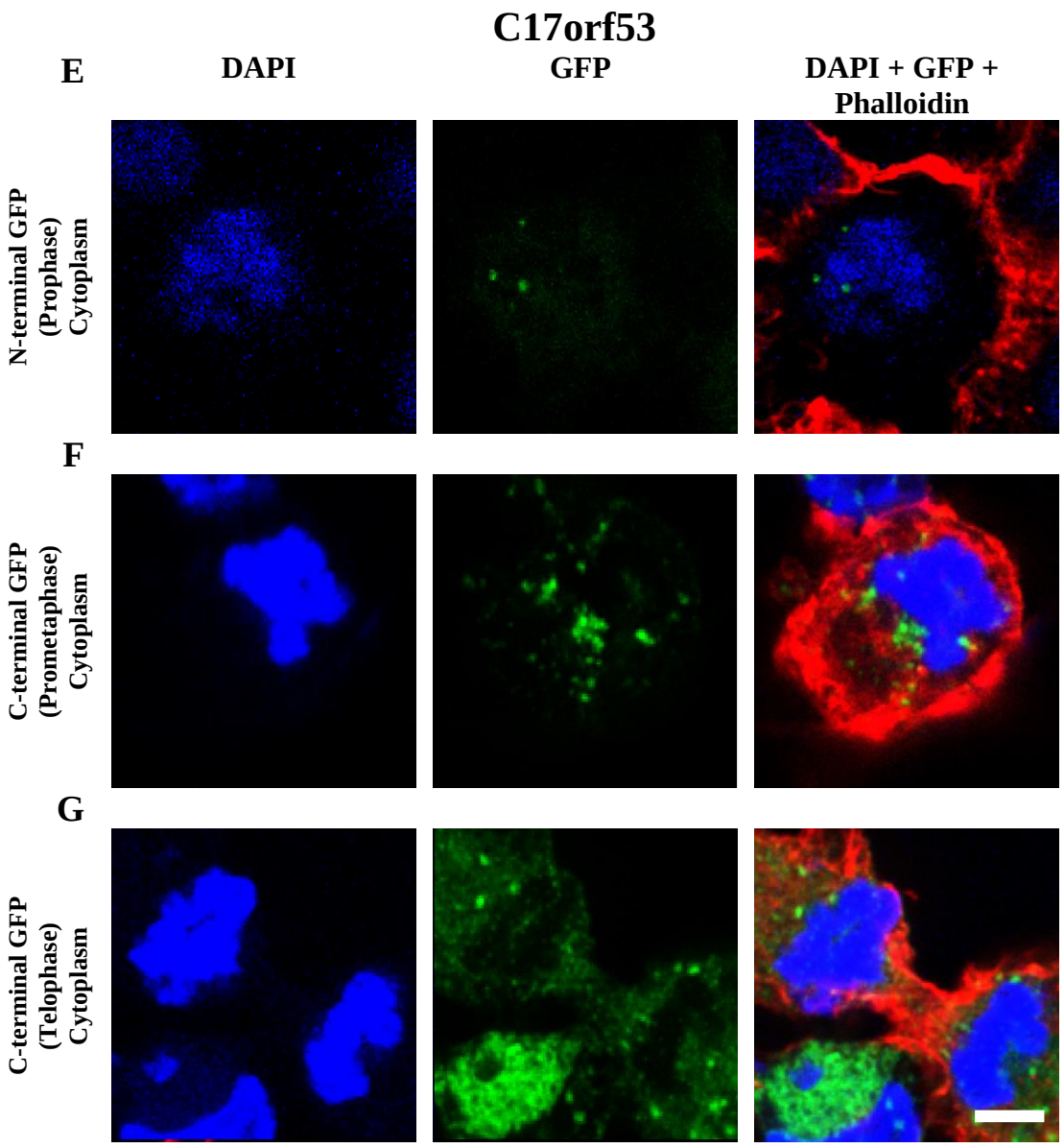


Figure 4.6: Subcellular localisation of C17orf53 in HEK293T cells. Expression of C17orf53 tagged with GFP at the C-terminal during interphase **(A** and **B)**, prometaphase **(F)**, and telophase **(G)**, respectively. Expression of C17orf53 tagged with GFP at the N-terminal during interphase **(C** and **D)** and prophase **(E)**. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

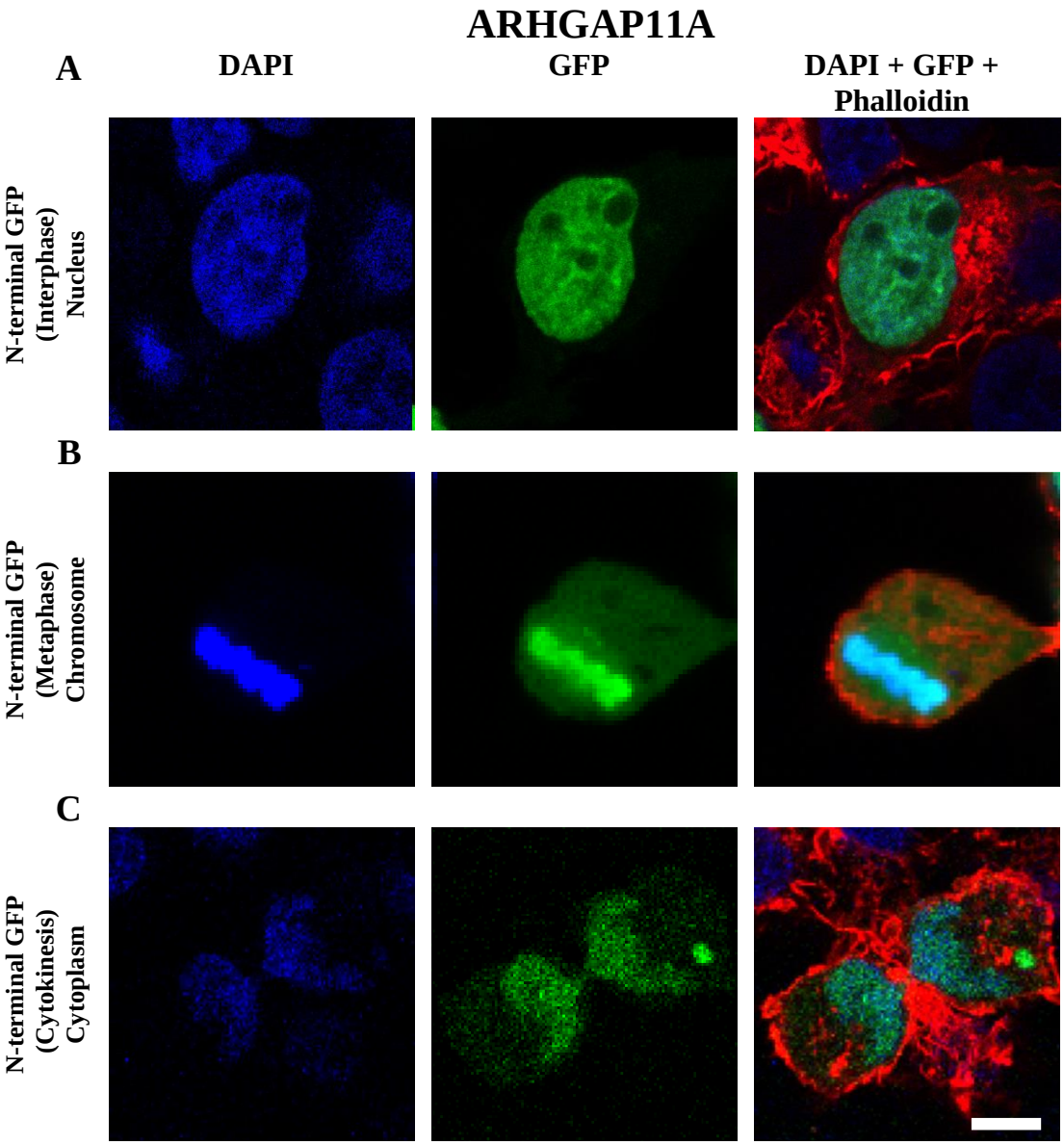


Figure 4.7: Subcellular localisation of ARHGAP11A in HEK293T cells. The localisation of ARHGAP11A tagged with GFP at the N-terminal expression during **A)** interphase, **B)** metaphase, and **C)** cytokinesis. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 6 µm.

FIGNL1 tagged with GFP at the N-terminal was expressed in HEK293T cells. In interphase, there is a single spot of expression proximal to the nucleus in the cytoplasm (Figure 4.8A). During prophase, there are two bright spots of GFP expression proximal to the nucleus with weaker granular expression throughout the cytoplasm (Figure 4.8B). In prometaphase, the majority of GFP expression appears to be proximal to the mitotic chromosomes (Figure 4.8C).

FAM111A tagged with GFP was expressed in HEK293T cells. In interphase and prophase, FAM111A with N-terminal GFP tagged protein is expressed with one or two spots proximal to the nucleus and weaker granular expression throughout the cytoplasm (Figures 4.9A and B). FAM111A tagged with GFP at the C-terminal is expressed as a cluster of spots in the nucleus and the cytoplasm during interphase (Figure 4.9C).

RIBC2 was expressed in HEK293T cells to show its localisation. In interphase, RIBC2 tagged with GFP at the N-terminal is expressed in two spots proximal to the nucleus in the cytoplasm (Figures 4.10A and B). These spots appear to move to opposite poles of the cell (Figure 4.10B). In prophase, RIBC2 is expressed as a number of spots proximal to the condensed chromosome (Figure 4.10C). During prometaphase, the majority of spots appear to be localised to opposite poles proximal to mitotic chromosomes with weaker granular expression throughout the cytoplasm (Figure 4.10D). During metaphase, the expression of RIBC2 appears to be a clump of spots associated with metaphase chromosomes (Figure 4.10E). In anaphase, the majority of RIBC2 expression appears to be localised to opposite poles proximal to anaphase chromosomes with weaker granular expression in the cytoplasm (Figure 4.10F).

Genes of unknown function identified by RNAi screening

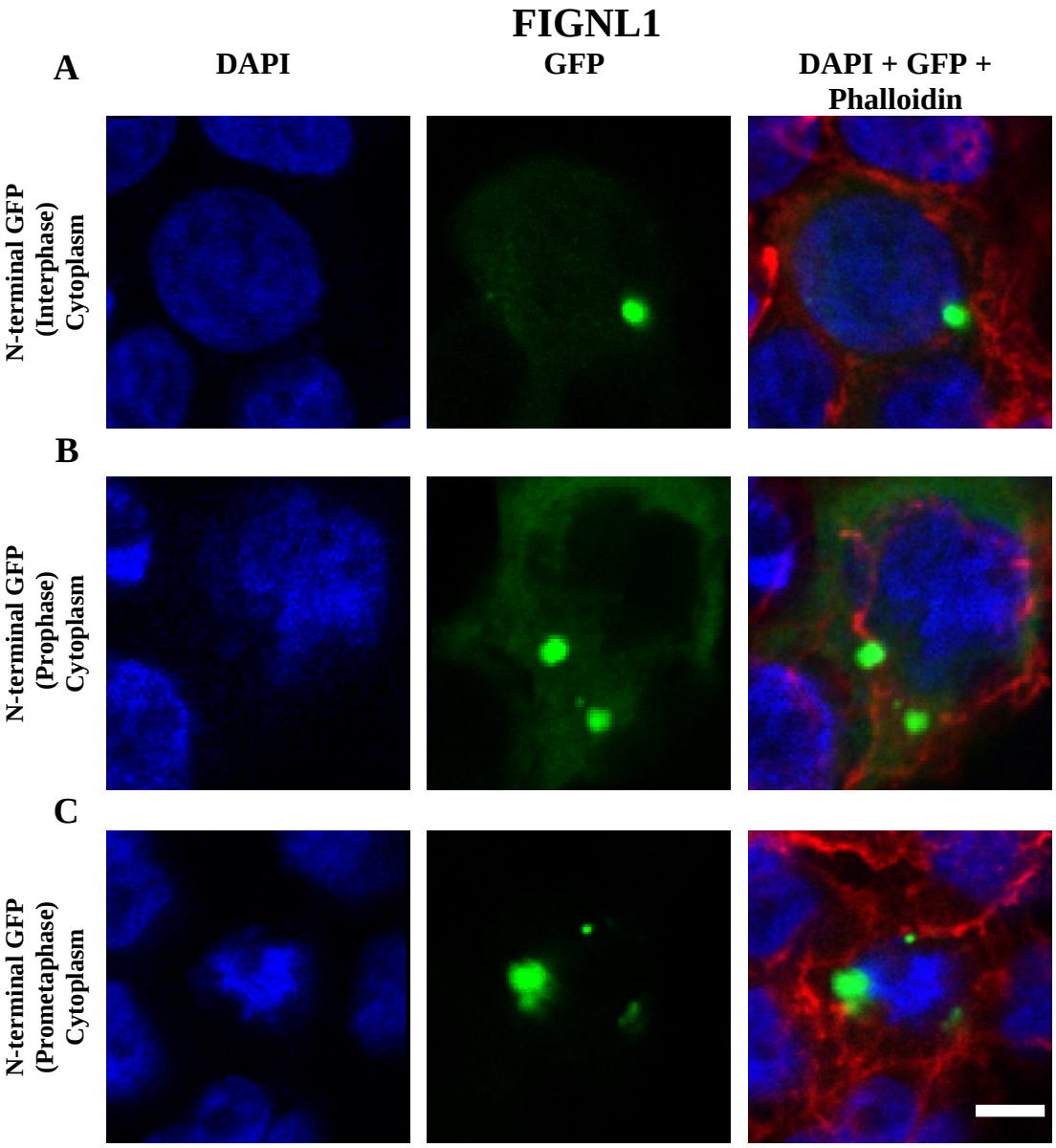


Figure 4.8: Subcellular localisation of FIGNL1 in HEK293T cells. The localisation of FIGNL1 expression during **A)** interphase, **B)** prophase, and **C)** metaphase. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

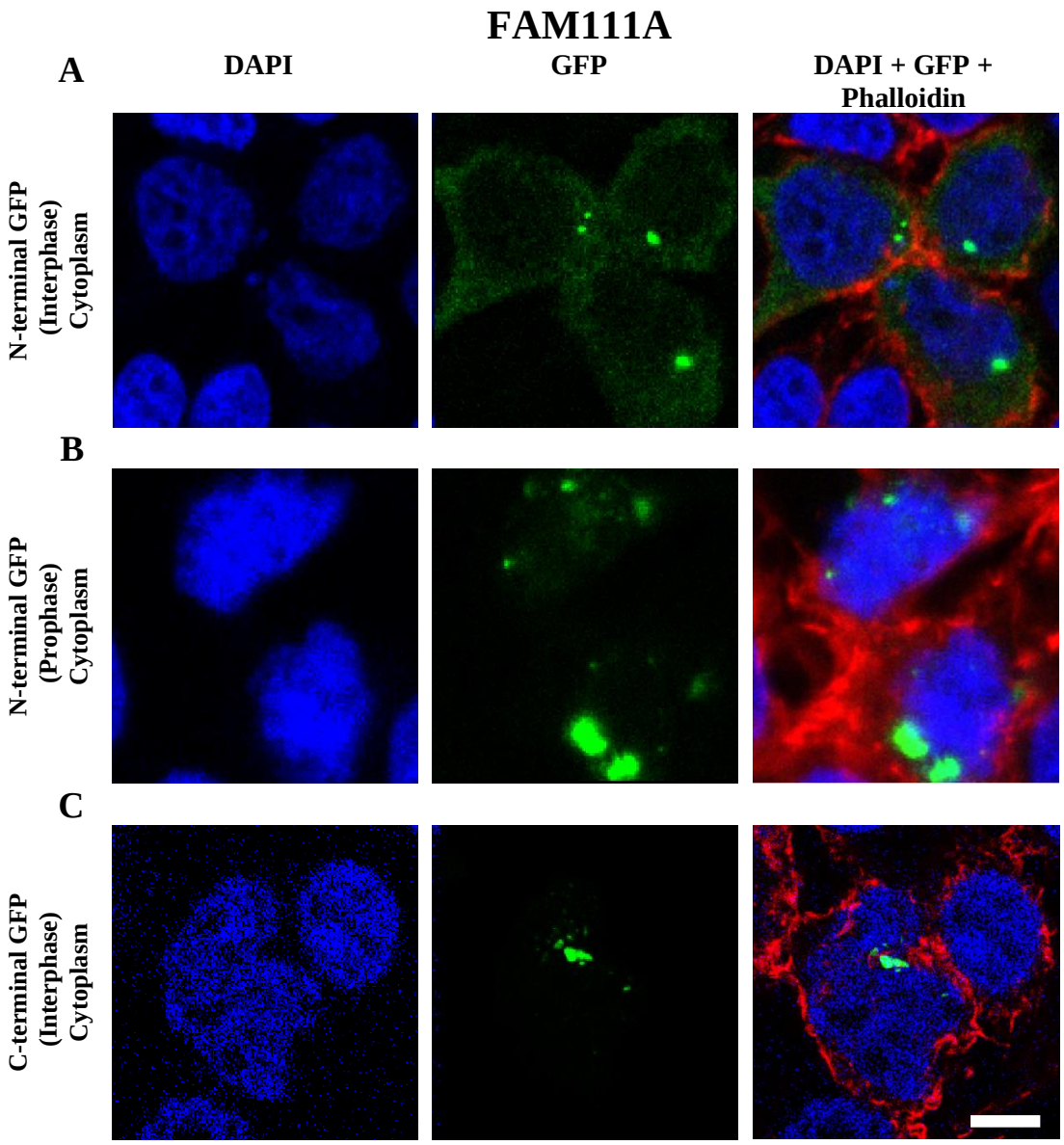
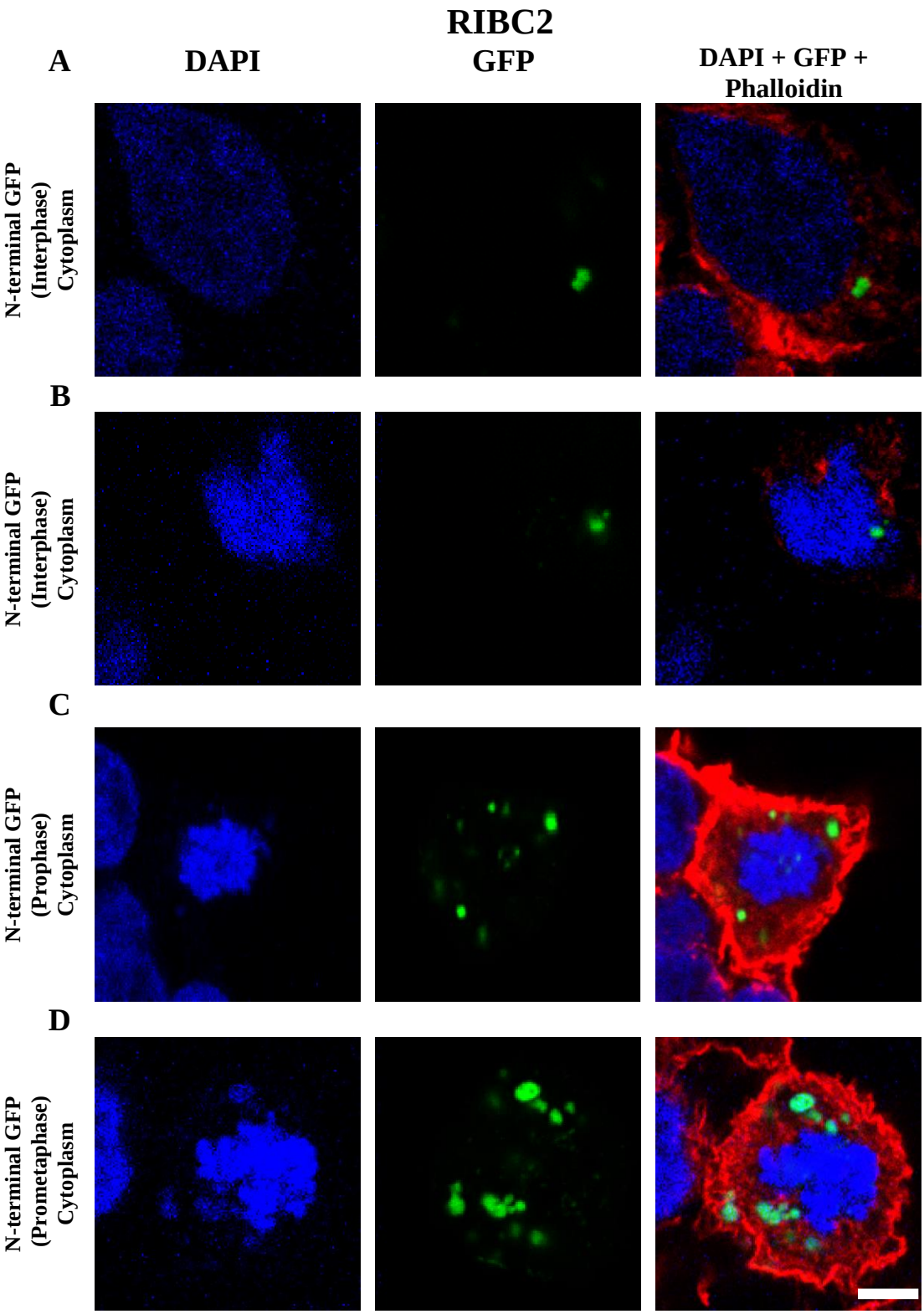


Figure 4.9: Subcellular localisation of FAM111A in HEK293T cells during interphase. A) and **B)** show the localisation of FAM111A expression tagged with GFP at the N-terminal during interphase. **C)** The localisation of FAM111A expression tagged with GFP at the C-terminal during interphase. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

Genes of unknown function identified by RNAi screening



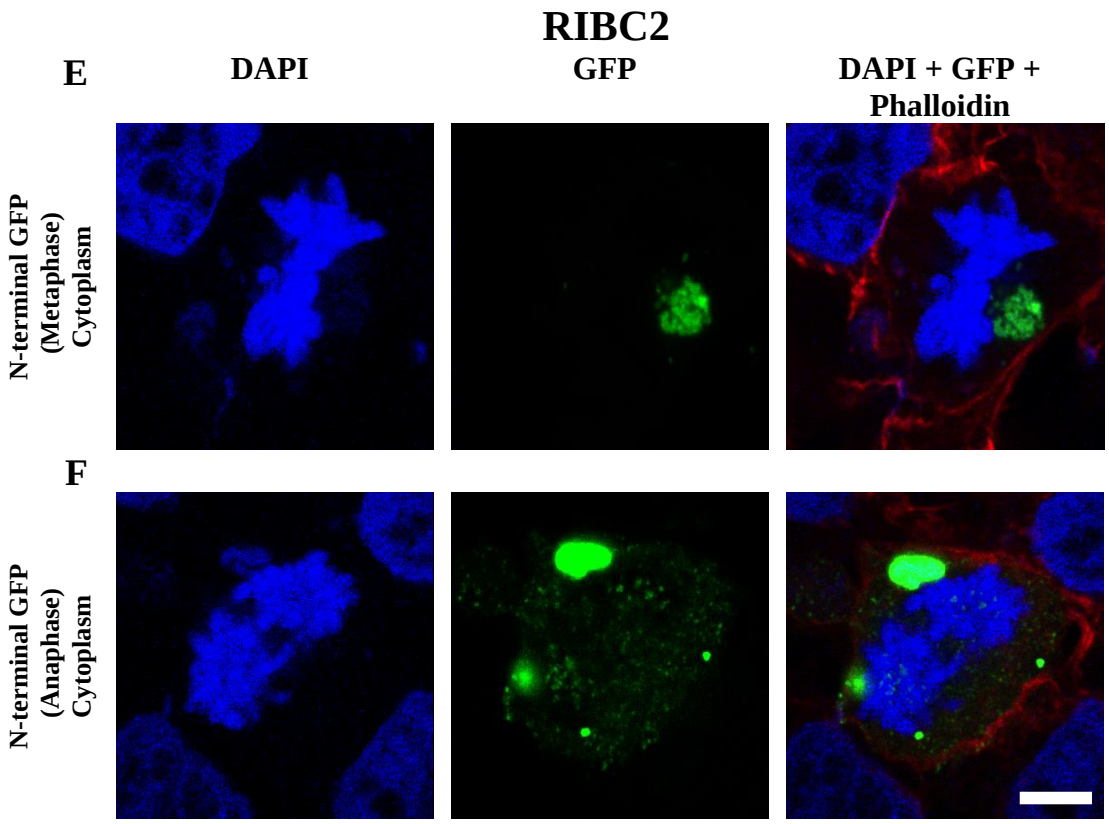


Figure 4.10: Subcellular localisation of RIBC2 tagged with GFP at the N-terminal in HEK293T cells. RIBC2 expression in the cytoplasm during **A)** interphase, **B)** prophase, **C)** prophase **D)** prometaphase, **E)** metaphase, and **F)** anaphase. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm confocal sections. Scale bar 4 µm.

During interphase, KIAA1549L tagged with GFP at the N-terminal is expressed as two spots proximal to the nucleus with weaker granular expression in the cytoplasm (Figure 4.11A). In contrast, expression of KIAA1549L tagged with GFP at the C-terminal shows weaker granular expression throughout the nucleus but not the nucleoli (Figure 4.11B).

The protein expression of RPL39L tagged with GFP at the N-terminal appears to be localised to the nucleus in interphase (Figure 4.12). A number of green spots are randomly positioned in the nucleus with weaker granular expression (Figure 4.12A, B, and C). To determine whether this is due to the overexpression of clones, the expression of RPL39L fused with GFP at the C-terminal is required.

REEP4 tagged with GFP was expressed in HEK293T cells. During interphase, the protein expression of REEP4 tagged with GFP at the N-terminal is localised throughout the cytoplasm (Figure 4.13A). In prophase, the spots of REEP4 expression tagged with GFP at the N-terminal are proximal to the chromosome with weaker granular expression in the cytoplasm (Figure 4.13B). For expression of REEP4 with GFP tagged at the C-terminal, two spots are proximal to the nucleus during interphase (Figure 4.13C). During prophase, the majority of spots are associated with the condensed chromosome with weaker granular expression in the cytoplasm (Figure 4.13D)

GSTCD was tagged with GFP at the N-terminal and expressed in HEK293T cells (Figure 4.14). Granular expression of GSTCD was found to be unevenly distributed throughout the cytoplasm during interphase (Figures 4.14A, B, and C). Additional localisation of GSTCD tagged with GFP at the C-terminal is required to determine whether or not the expression of GSTD is an artefact.

UBR7 tagged with GFP at the N-terminal was expressed in HEK293T cells (Figure 4.15). The granular expression of UBR7 appears to be ubiquitously localised to the nucleus but not nucleoli in interphase. To determine whether the pattern of UBR7 expression is an artefact, further subcellular protein localisation of UBR7 tagged with GFP at the C-terminal is required.

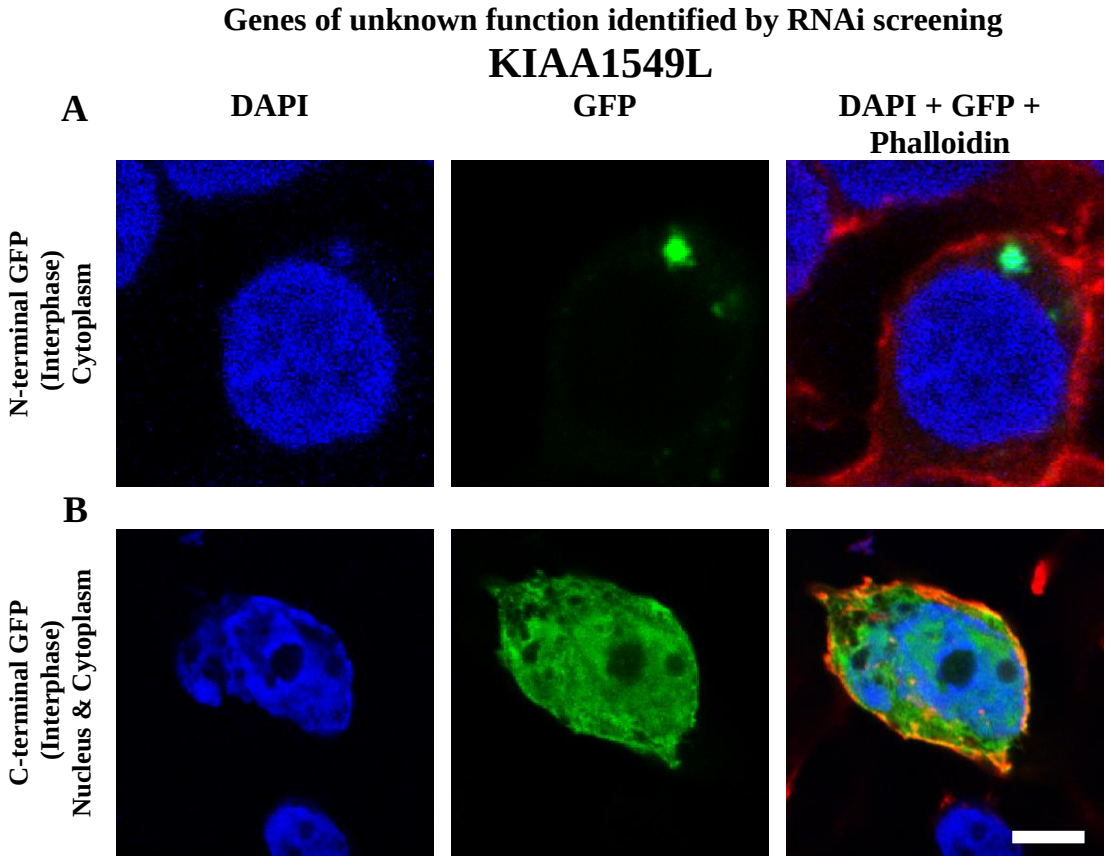


Figure 4.11: Subcellular localisation of KIAA1549L in HEK293T cells. A) Expression KIAA1549L tagged with GFP at the N-terminal during interphase in the cytoplasm. **B)** Expression of KIAA1549L tagged with GFP at the C-terminal in the nucleus and cytoplasm during interphase**.** Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

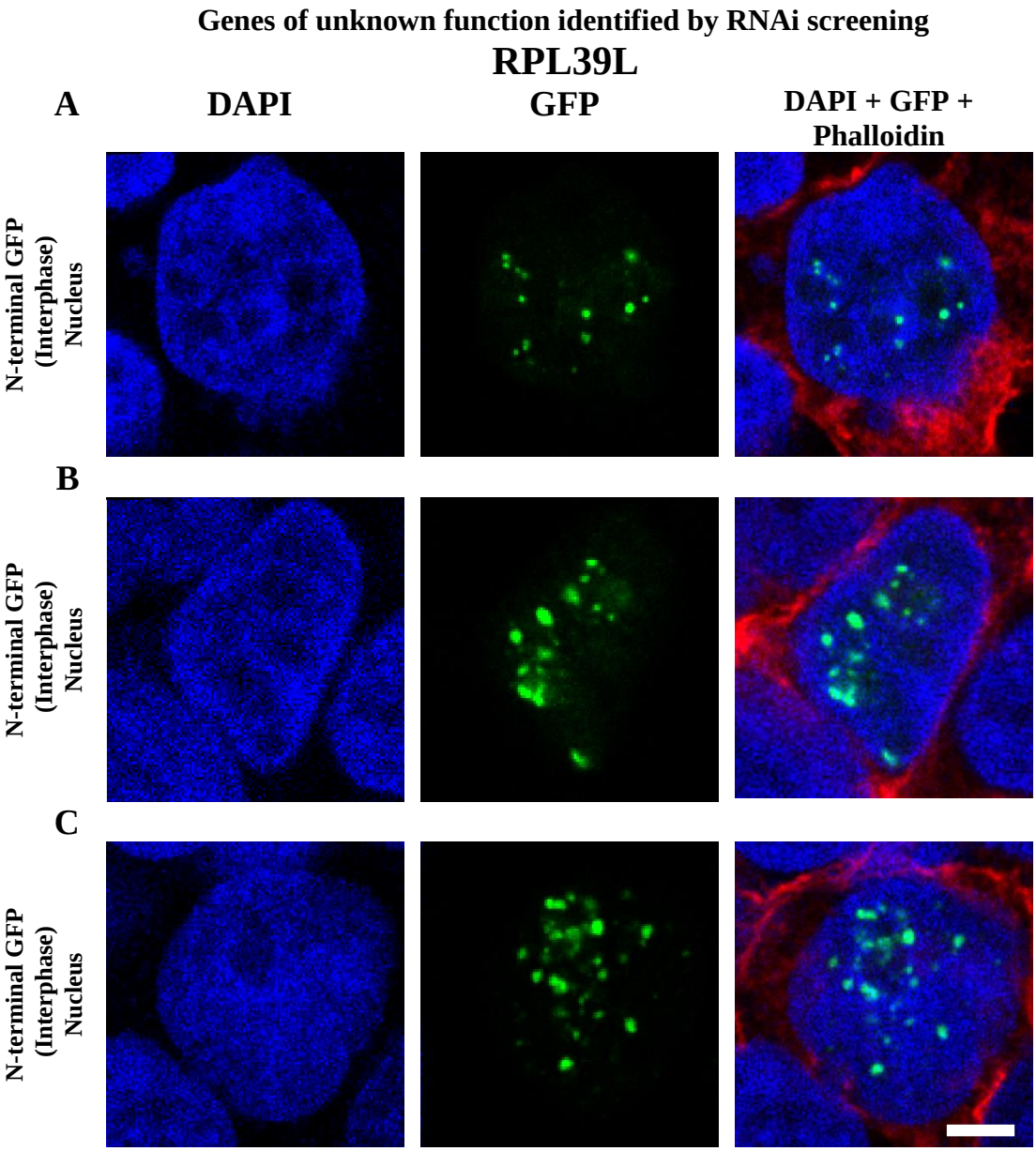


Figure 4.12: Subcellular localisation of RPL39L in HEK293T cells. A) B) and **C)** Expression of RPL39L expression tagged with GFP at the N-terminal is localised to the nucleus during interphase. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

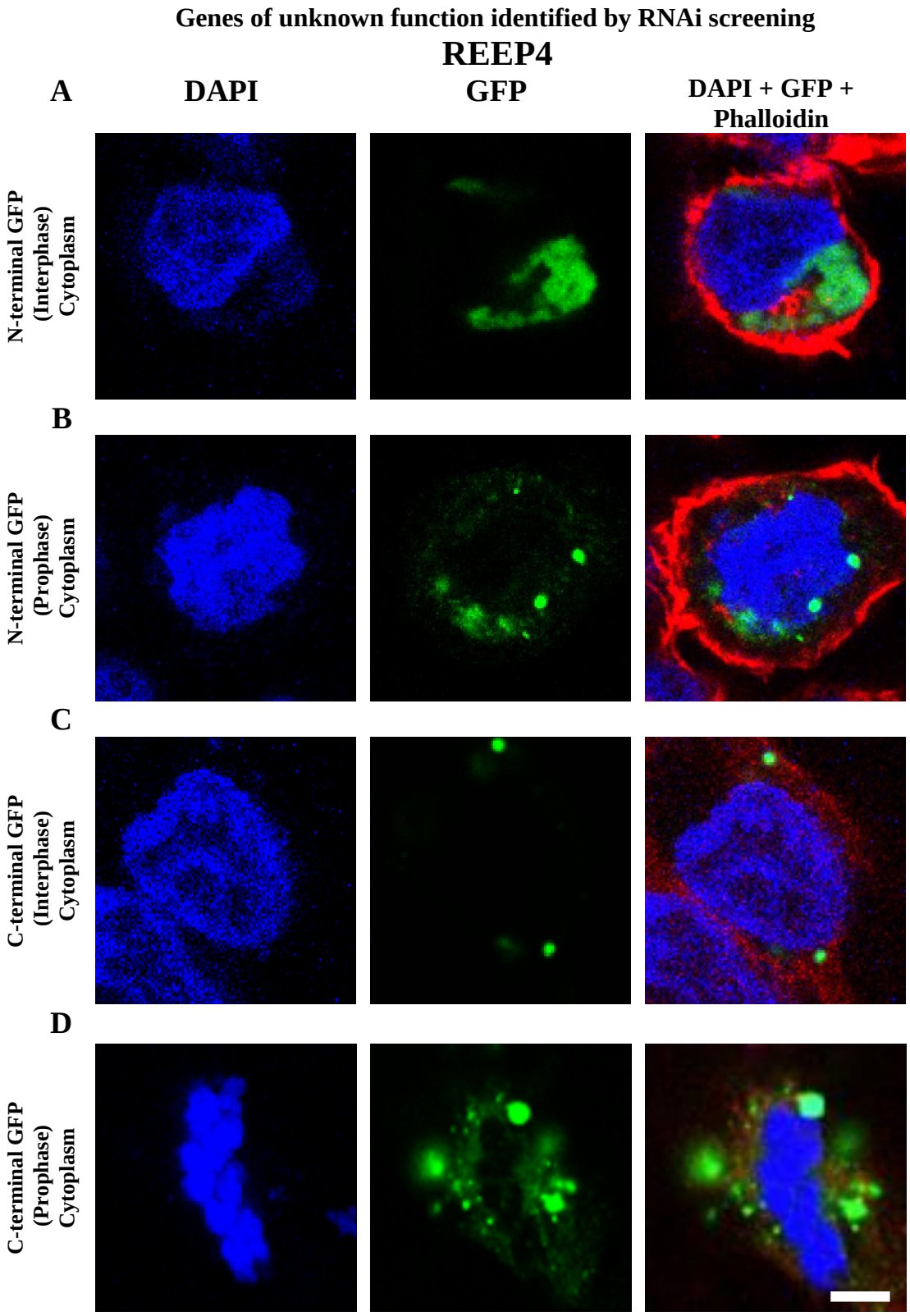


Figure 4.13: Subcellular localisation of REEP4 in HEK293T cells. The localisation of REEP4 expression during **A** and **B)** interphase and **C** and **D)** prophase, respectively. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

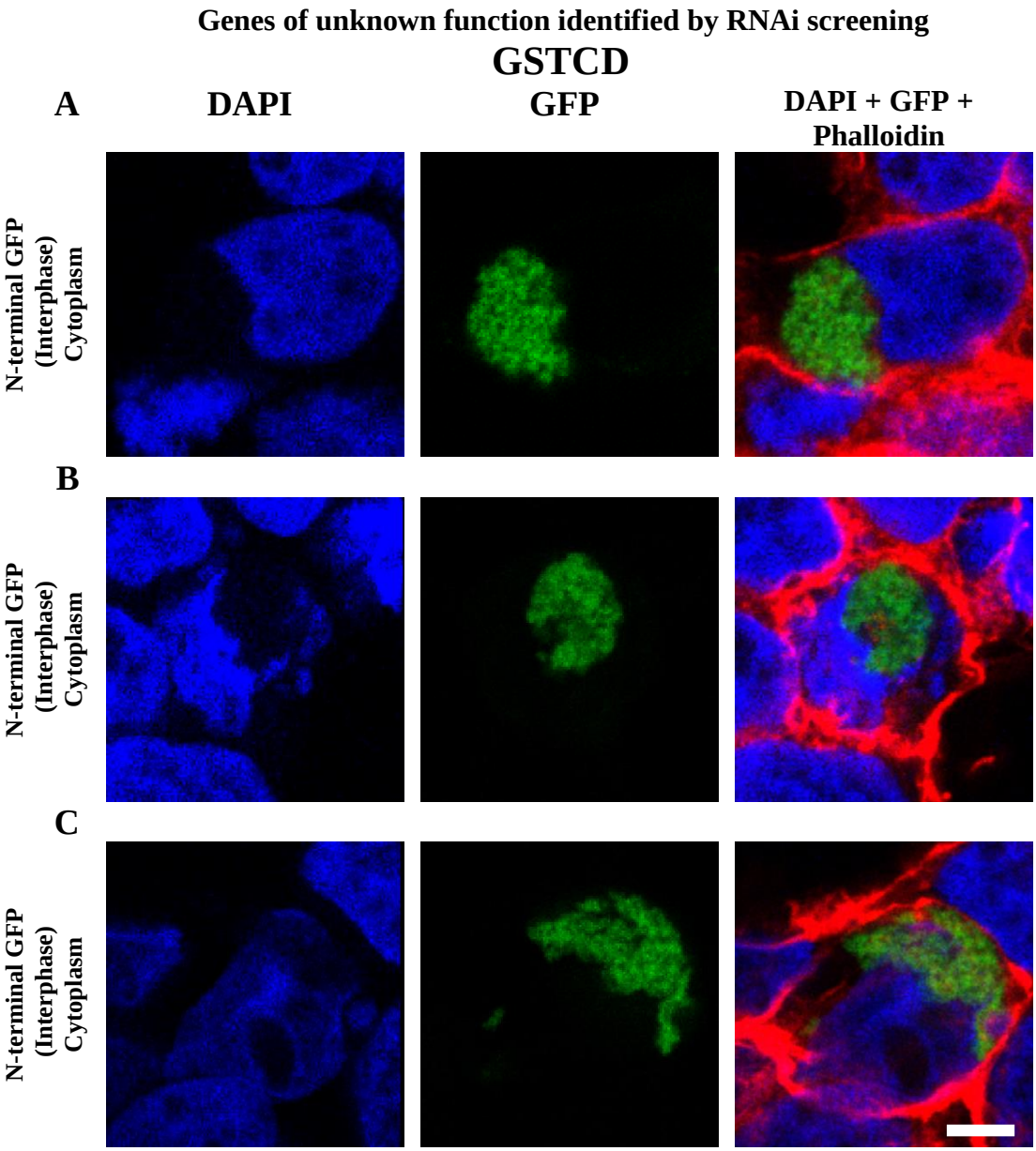


Figure 4.14: Subcellular localisation of GSTCD in HEK293T cells. A), B) and **C)** Expression of GSTCD expression tagged with GFP at the N-terminal is localised to the cytoplasm. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X are 1.6 µm optical sections. Scale bar 10 µm.

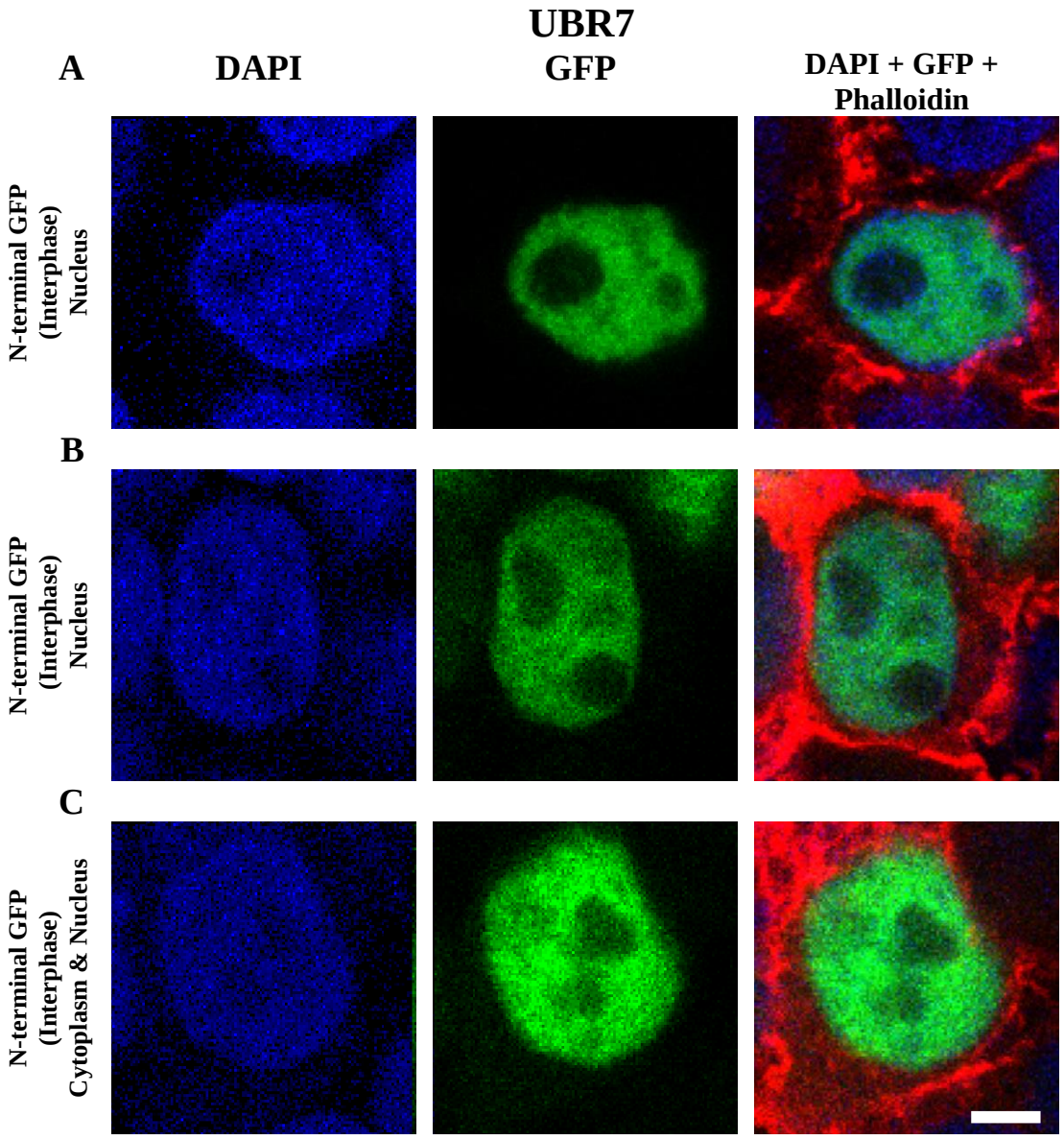


Figure 4.15: Subcellular localisation of UBR7 in HEK293T cells. A), B), and **C)** Expression of UBR7 tagged with GFP at the N-terminal is localised to the nucleus but not the nucleolus. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X are 1.6 µm optical sections. Scale bar 4 µm.

DEPDC1B tagged with GFP was expressed in HEK293T cells. The expression of DEPDC1B with GFP at the N-terminal shows a number of clusters of spots with weaker granular expression throughout the nucleus (Figure 4.16A). The expression of DEPDC1B tagged with GFP at the C-terminal shows granular expression throughout the cytoplasm and nucleus (Figures 4.16B and C).

ANLN tagged with GFP at the N-terminal was expressed in HEK293T cells. Expression of ANLN during interphase was found throughout the cytoplasm but appeared uneven in distribution (Figures 4.17A, B, and C).

DEPDC1 tagged with GFP at the N-terminal was expressed in HEK293T cells. DEPDC1 was expressed as a clump of spots in the nucleoli and cytoplasm during interphase (Figures 4.18A and C). Spots as can be seen proximal to the cell membrane associated with weaker granular expression in interphase (Figure 4.18B).

FAM111B tagged with GFP was expressed in HEK293T cells. FAM111B tagged with GFP at the N-terminal was expressed as one spot proximal to the nucleus during interphase (Figure 4.19A). FAM111B tagged with GFP at the C-terminal was expressed as two spots proximal to the nucleus during interphase (Figure 4.19B). It is interesting that the expression of FAM111B tagged with GFP at the C-terminal shows two spots proximal to the nucleus but at opposite poles (Figure 4.19C).

KIAA0101 tagged with GFP was expressed in HEK293T cells. The localisation pattern of KIAA0101 tagged with GFP at the N-terminal exhibited a thick dashed line (worm like) throughout the nucleus. Expression of KIAA0101 tagged with GFP at the C-terminal exhibits spots and clumps throughout the nucleus (Figure 4.20C).

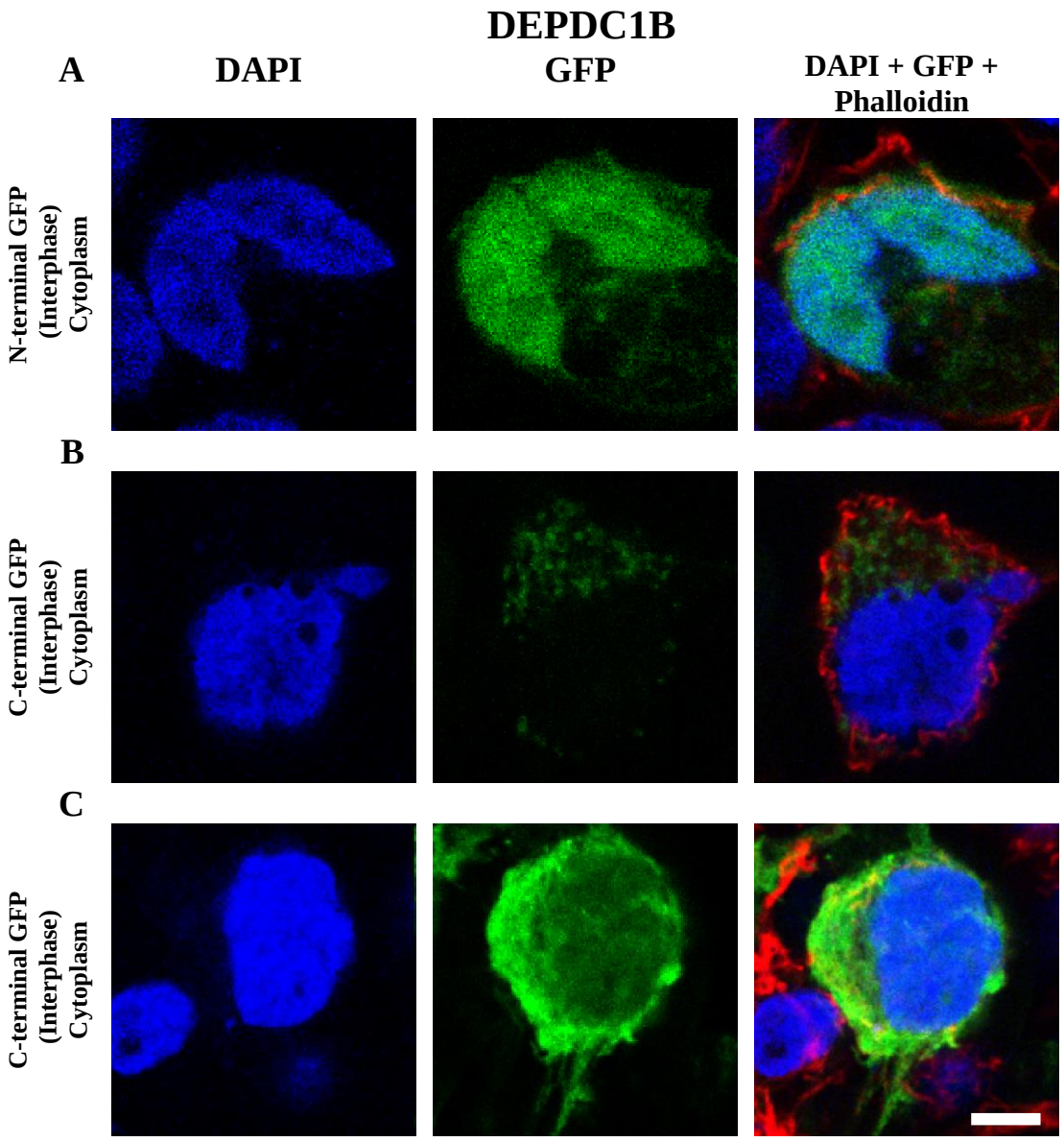


Figure 4.16: Subcellular localisation of DEPDC1B in HEK293T cells. A) Expression of DEPDC1B expression tagged with GFP at the N-terminal was localised to the cytoplasm and nucleus. **B)** and **C)** Expression DEPDC1B tagged with GFP at the C-terminal was localised to the nucleus and cytoplasm. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X are 1.6 µm optical sections. Scale bar 4 µm.

Genes of unknown function identified by network structure analysis only

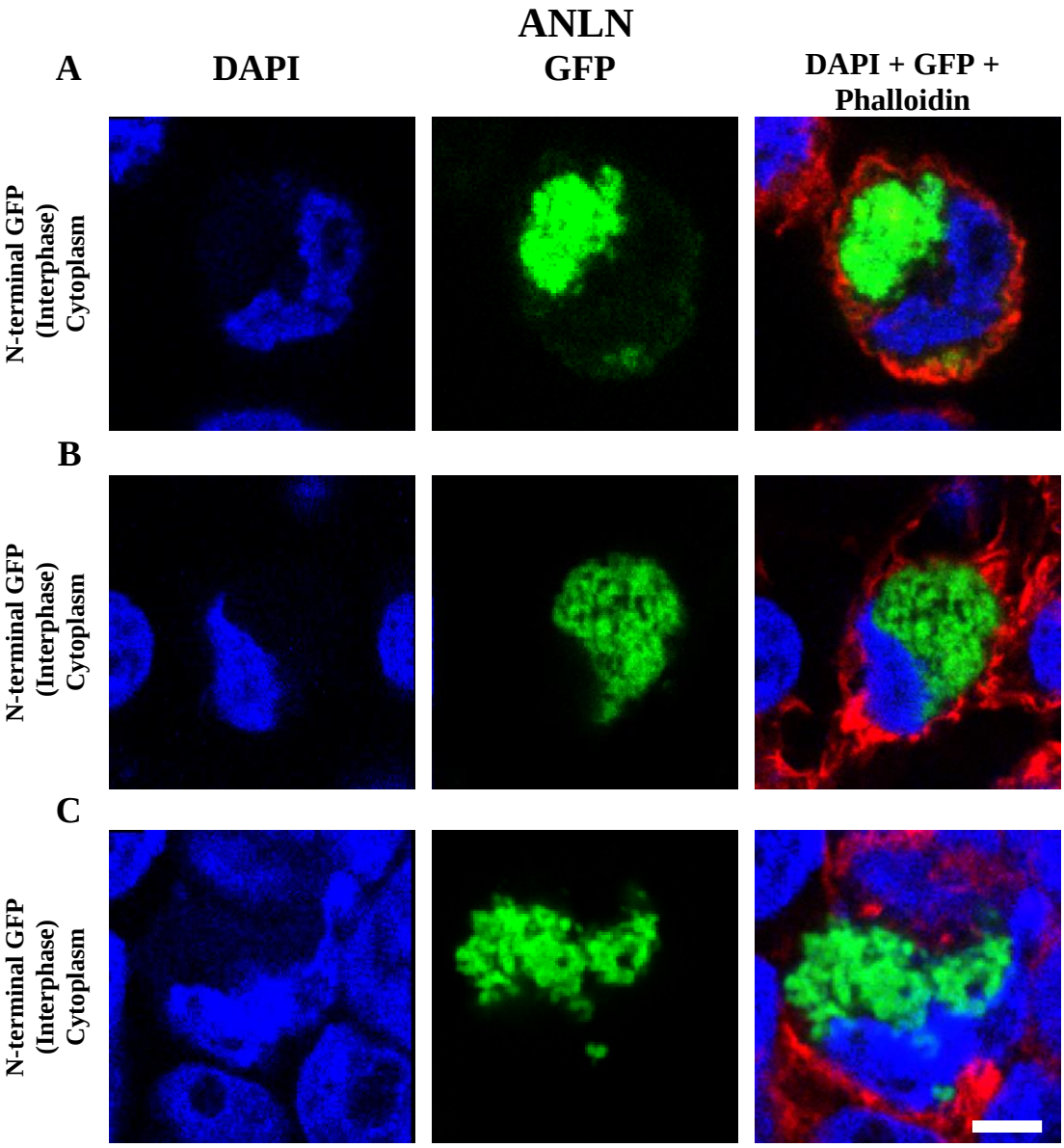


Figure 4.17: Subcellular localisation of ANLN in HEK293T cells. A), B) and **C)** Expression of ANLN expression tagged with GFP at the N-terminal was localised to the cytoplasm during interphase. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X are 1.6 µm optical sections. Scale bar 4 µm.

Genes of unknown function identified by network structure analysis only

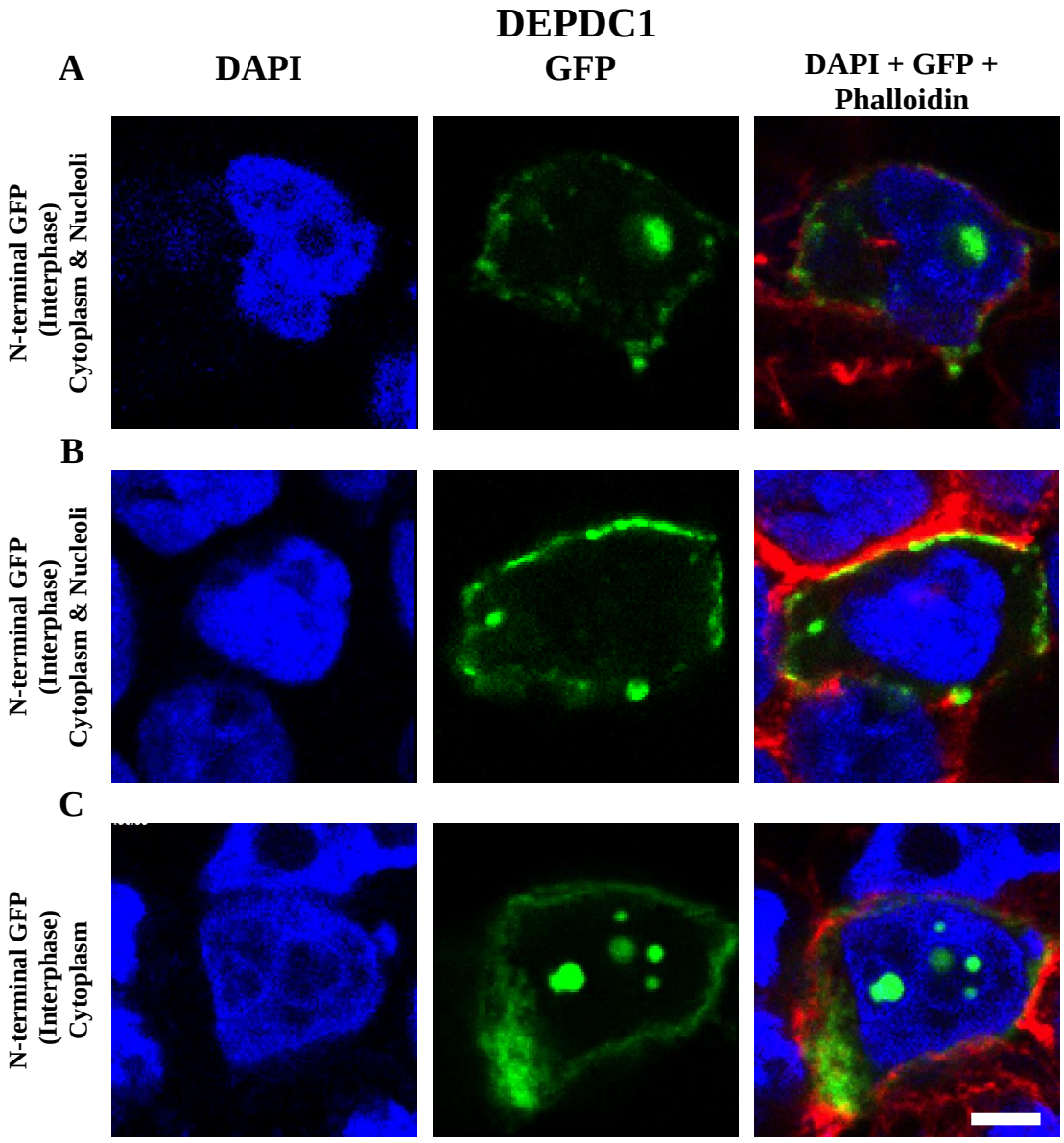


Figure 4.18: Subcellular localisation of DEPDC1 in HEK293T cells. A) and **B)** Expression of DEPDC1 tagged with GFP at the N-terminal was localised to cytoplasm and the nucleolus. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X are 1.6 µm optical sections. Scale bar 4 µm.

Genes of unknown function identified by network structure analysis only

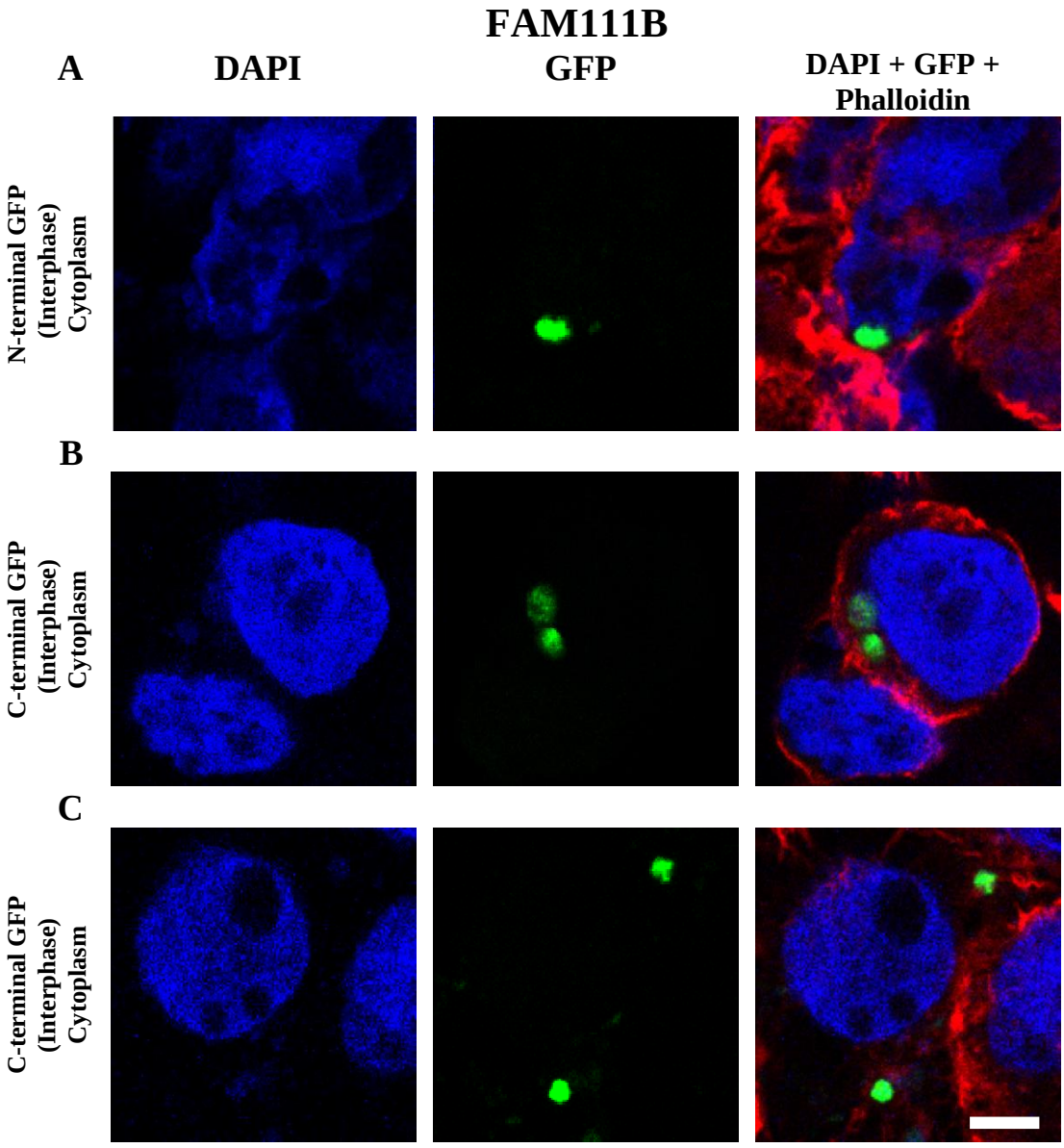


Figure 4.19: Subcellular localisation of FAM111B in HEK293T cells. A) Expression of FAM111B tagged with GFP at the N-terminal was localised at the cytoplasm during interphase. **B)** and **C)** The expression of FAM111B tagged with GFP at the C-terminal was localised to the cytoplasm during interphase. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X are 1.6 µm optical sections. Scale bar 4 µm.

Genes of unknown function identified by network structure analysis only

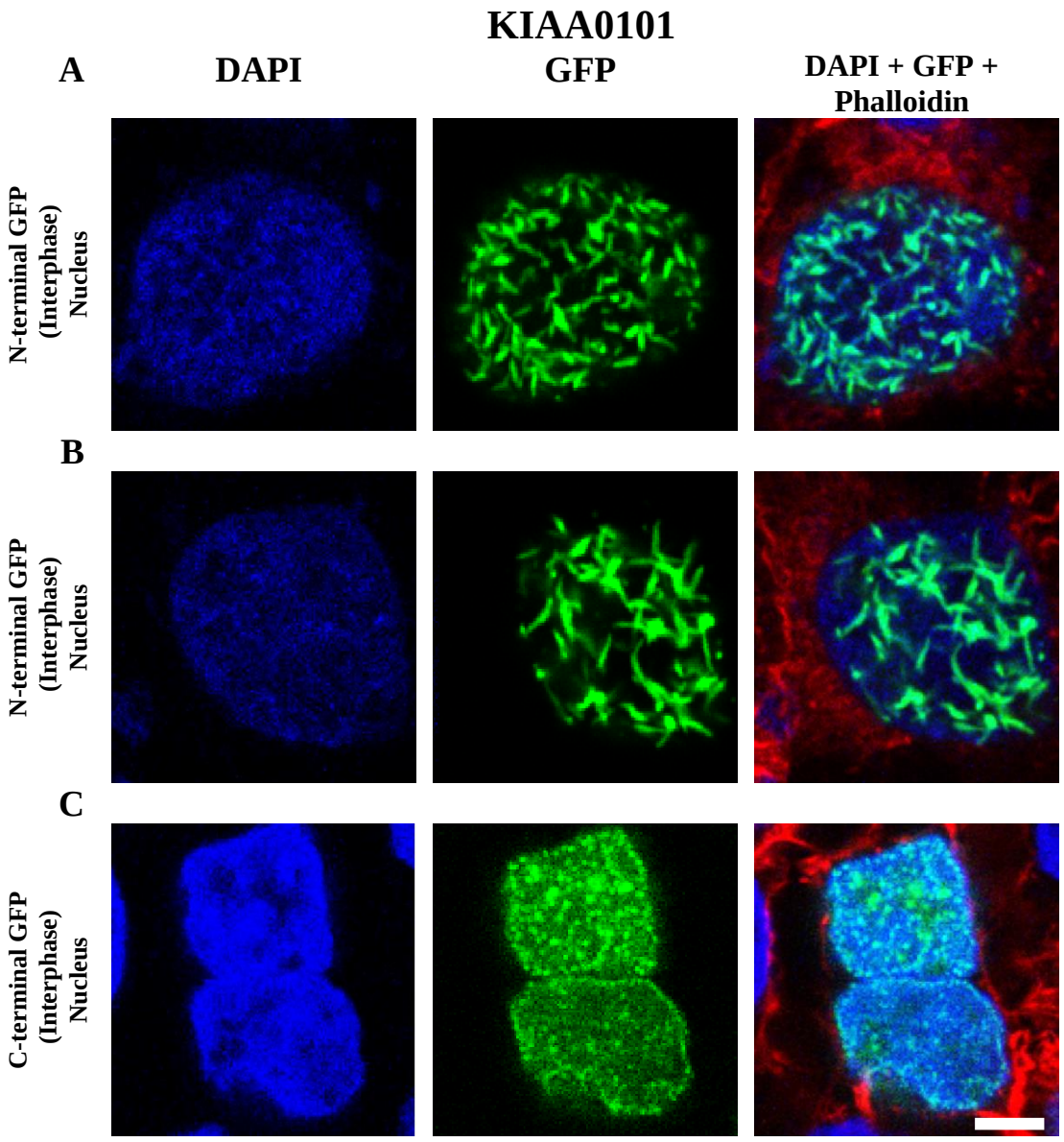


Figure 4.20: Subcellular localisation of KIAA0101 in HEK293T cells. A) and **B)** Expression of KIAA0101 tagged with GFP at the N-terminal was localised to the nucleus. **C)** Expression of KIAA0101 tagged with GFP at the C-terminal was localised to the nucleus. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

KNSTRN tagged with GFP at the N-terminal was expressed in HEK293T cells. KNSTRN was expression as a cluster of spots proximal to the nucleus in the cytoplasm (Figures 4.21A, B, and C).

KPNA2 tagged with GFP at the N-terminal was expressed in HEK293T cells. Expression of KPNA2 exhibited a discrete single spot with weaker granular expression proximal to the nucleus during interphase (Figures 4.22A and B).

RACGAP1P tagged with GFP at the N-terminal was expressed in HEK293T cells. During interphase, the expression of RACGAP1P exhibited a number of intense spots with weaker granular expression throughout the cytoplasm (Figure 4.23A). During prometaphase, the expression of RACGAP1P exhibited a number of clumps, spots with weaker granular expression in the cytoplasm (Figure 4.23B).

TCF19 tagged with GFP at the N-terminal was expressed in HEK293T cells. Expression of TCF19 exhibited a number of discrete spots with weaker granular expression throughout the nucleus but not the nucleoli (Figures 4.24A, B, and C).

Genes of unknown function identified by network structure analysis only

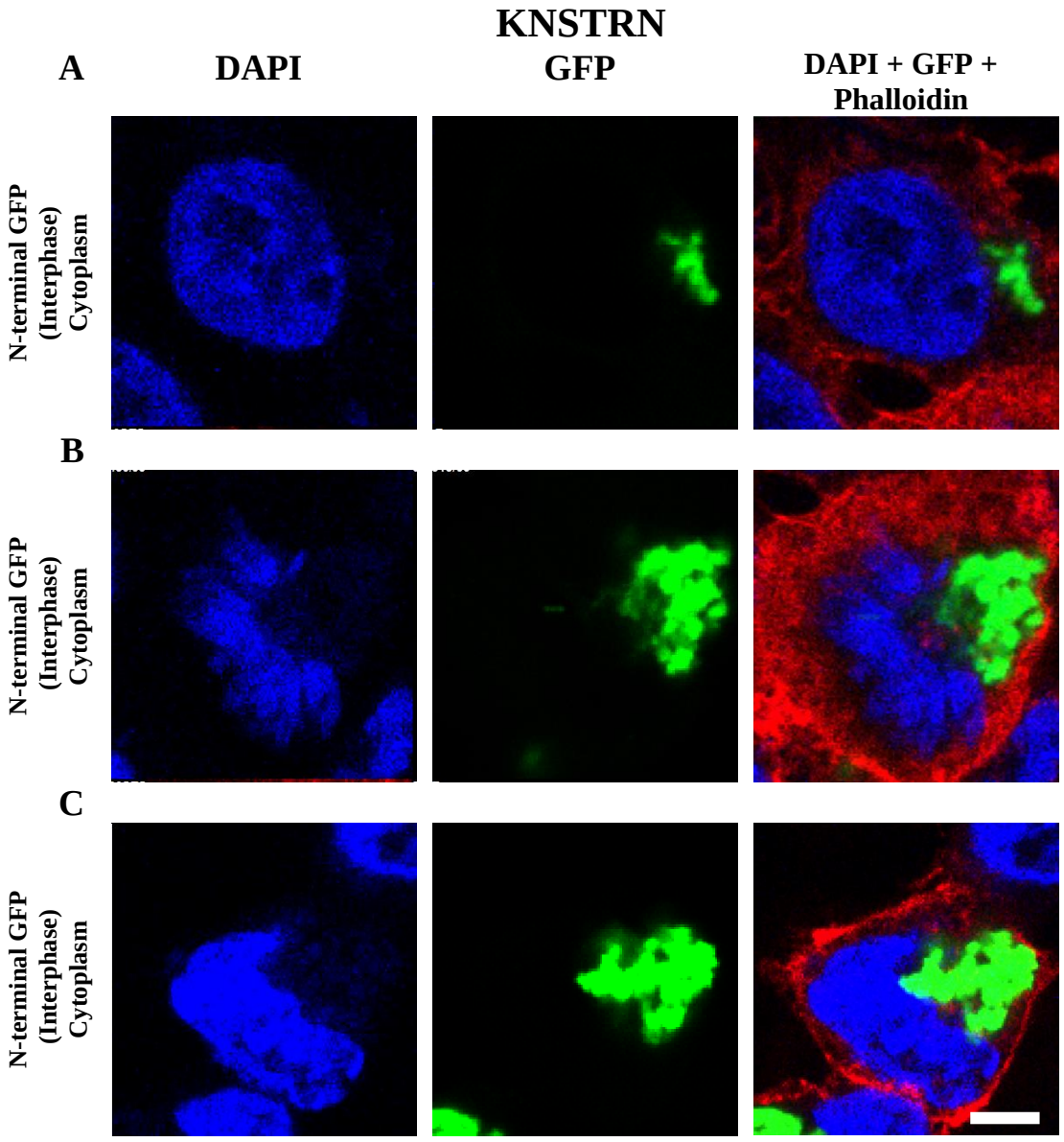


Figure 4.21: Subcellular localisation of KNSTRN in HEK293T cells. A), B), and **C)** Expression of KNSTRN tagged with GFP at the N-terminal was localised to the cytoplasm during interphase. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

Genes of unknown function identified by network structure analysis only

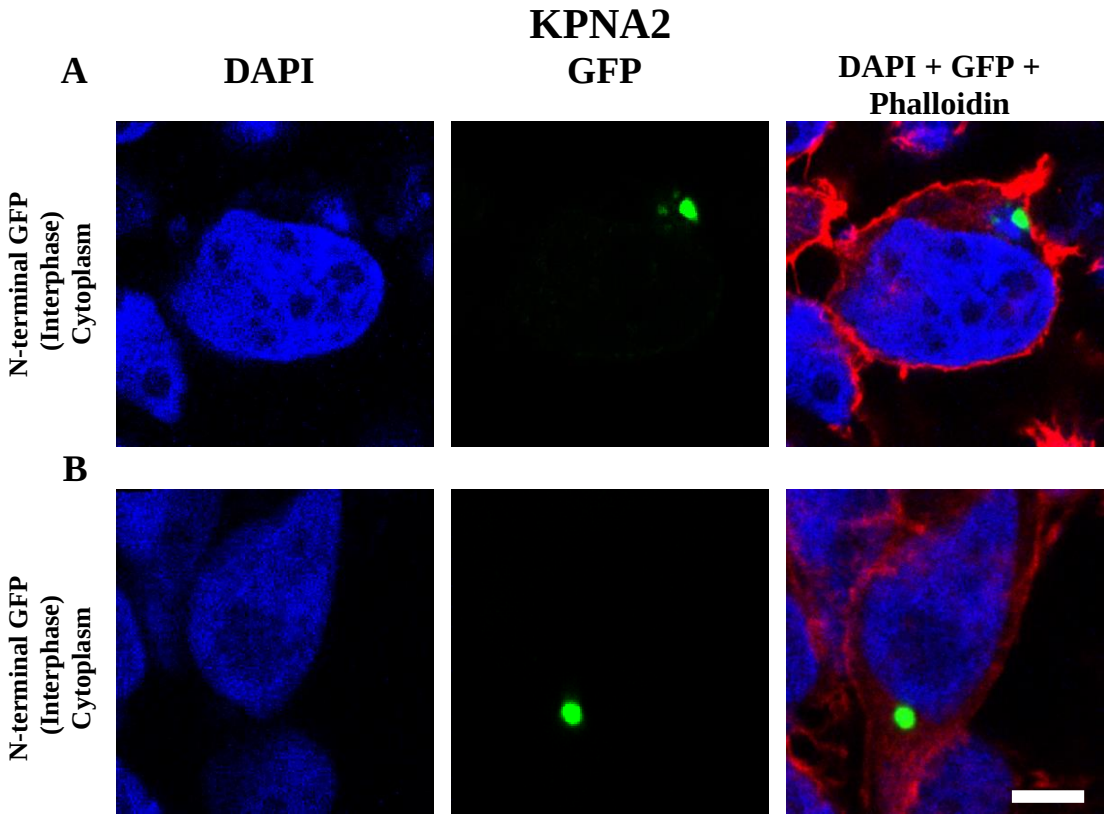


Figure 4.22: Subcellular localisation of KPNA2 in HEK293T cells. A) and **B)** Expression of KPNA2 tagged with GFP at the N-terminal was localised to the cytoplasm. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

Genes of unknown function identified by network structure analysis only

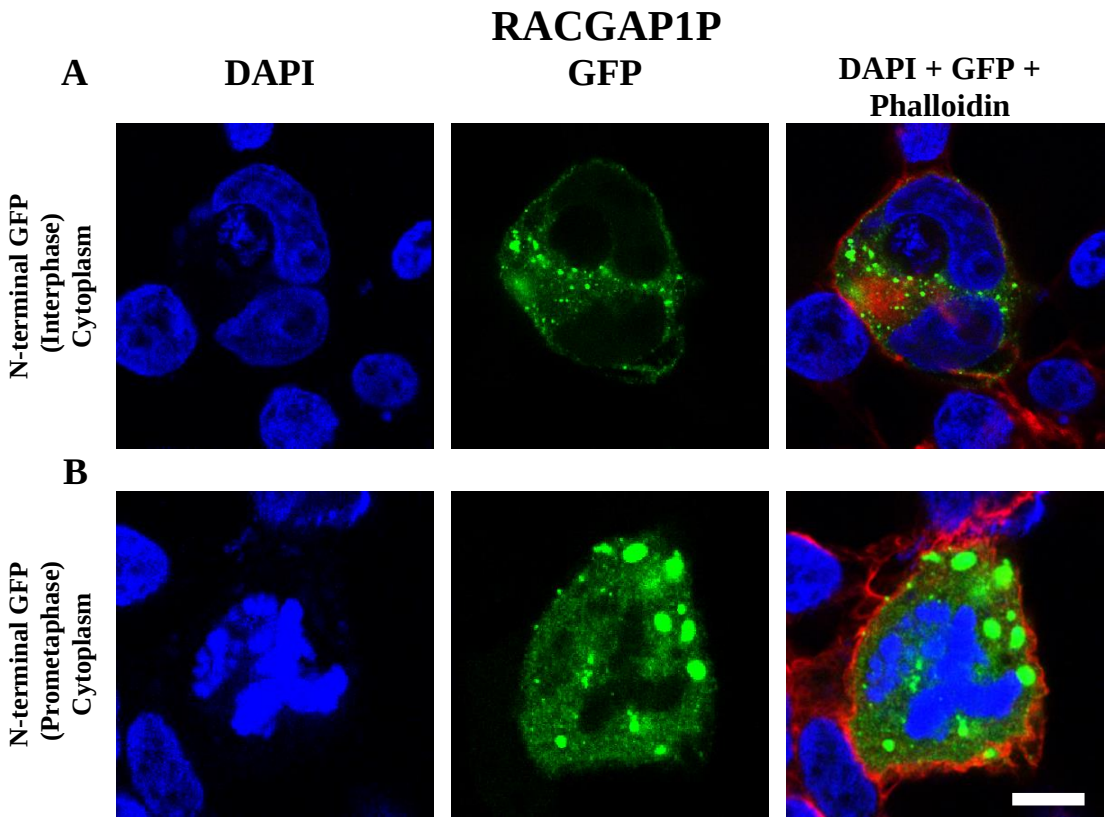


Figure 4.23: Subcellular localisation of RACGAP1P in HEK293T cells. A) and **B)** Expression of RACGAP1P tagged with GFP at the N-terminal was localised to the cytoplasm during interphase and prometaphase, respectively. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

Genes of unknown function identified by network structure analysis only

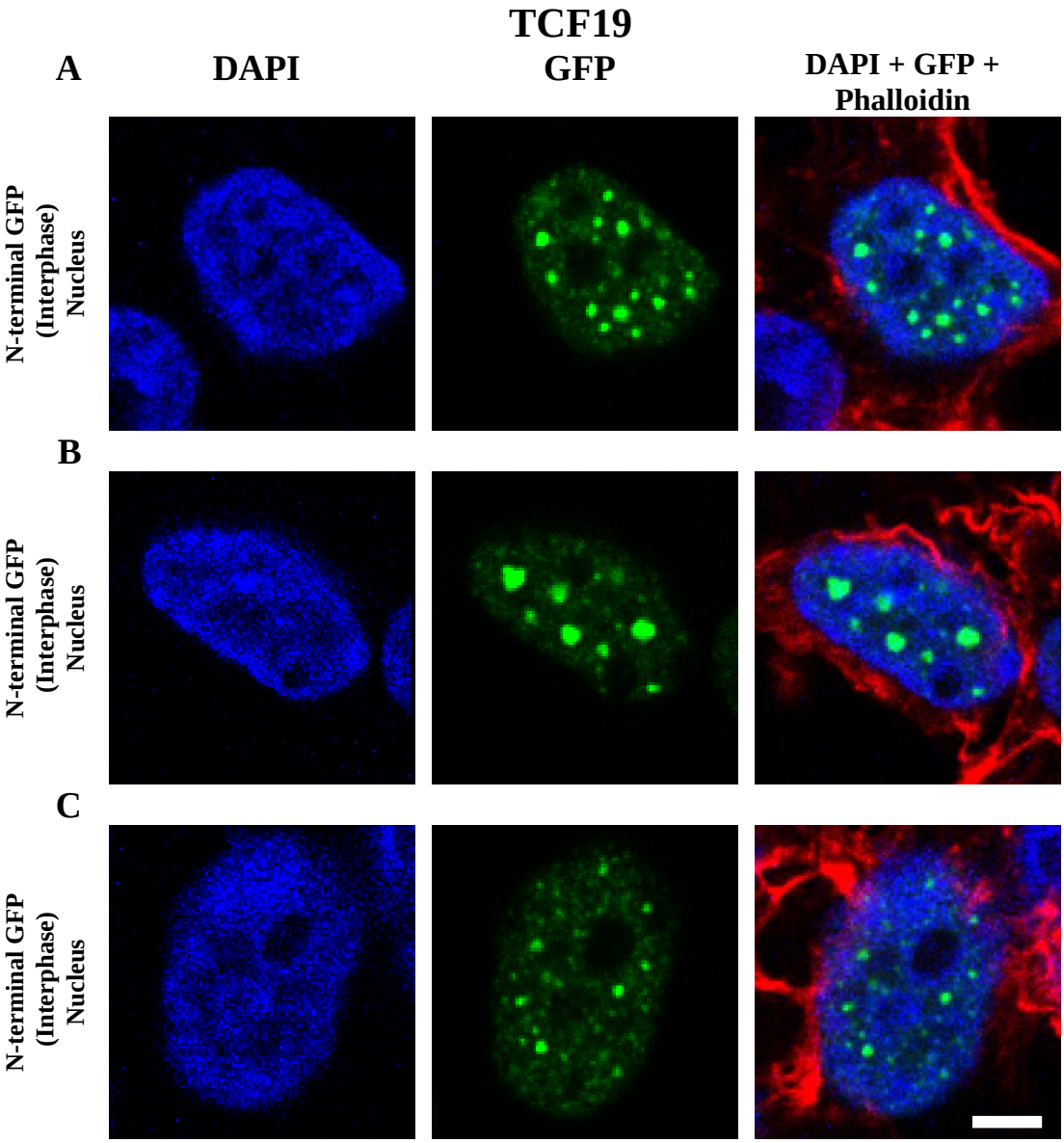


Figure 4.24: Subcellular localisation of TCF19 in HEK293T cells. A) and **B)** Expression of TCF19 tagged with GFP at the N-terminal was localised to the nucleus. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

To better understand expression of known cell cycle-regulated proteins within the centrosome, CEP55 expression tagged with GFP was used in this analysis. CEP55 tagged with GFP fusion at the N-terminal was expressed in HEK293T cells. During interphase, expression of CEP55 exhibited a cluster of spots proximal to the nucleus (Figure 4.25A). In prophase, the expression of CEP55 showed two clusters of spots proximal to the nucleus (Figure 4.25B). In prometaphase, the expression of CEP55 exhibited two or three spots with weaker granular expression throughout the cell with the exception of the mitotic chromosome (Figures 4.25C and D). In metaphase, CEP55 was expressed as two spots and another cluster of spots proximal to the mitotic chromosomes (Figure 4.25E). In anaphase, CEP55 was expressed as five spots and a large cluster of spots proximal to the sister chromatids (Figure 4.25F). The expression of CEP55 provides a good example to study the developing features of the centrosome in this analysis.

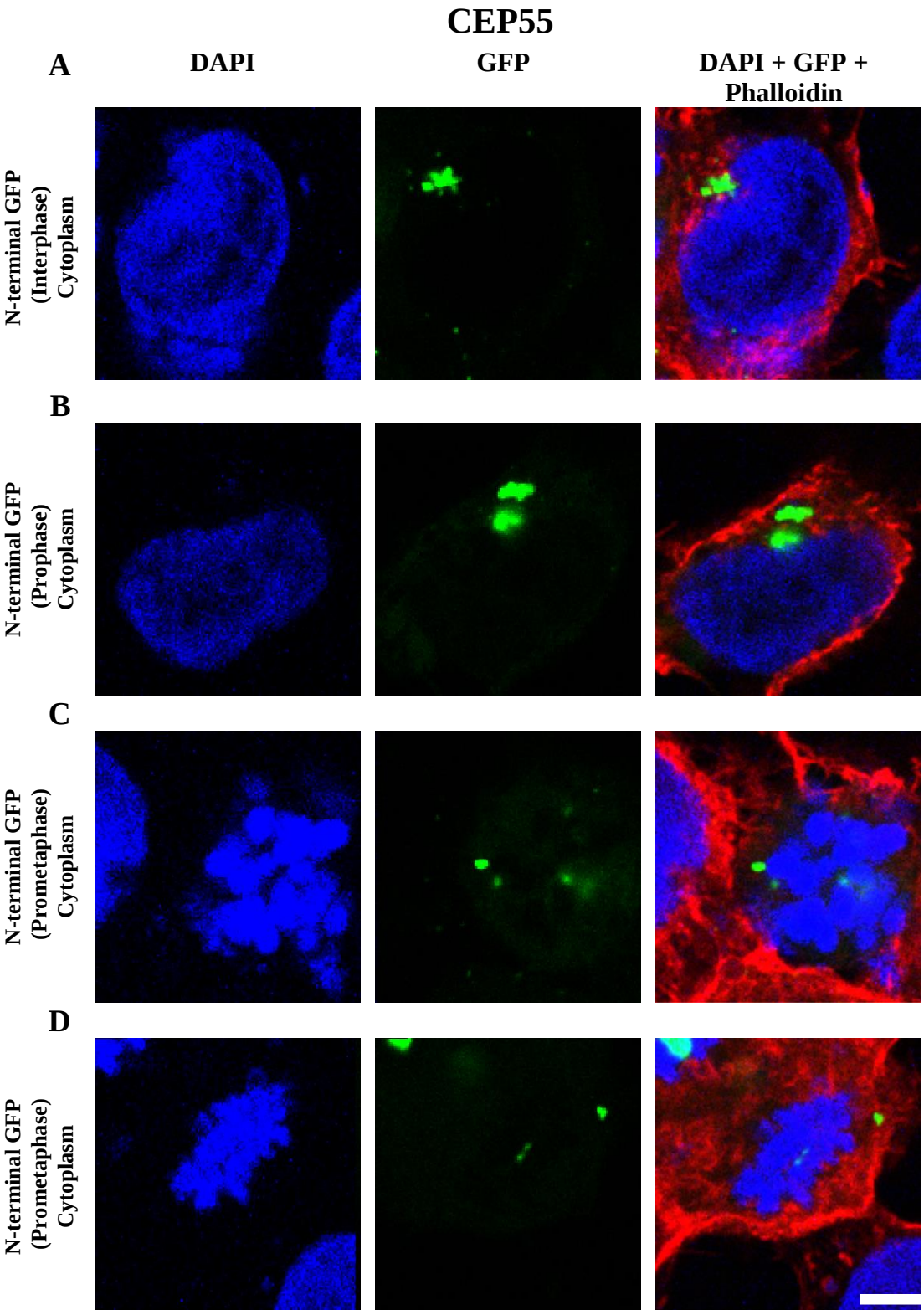
CEP57 tagged with GFP fusion at the N-terminal was expressed in HEK293T cells. During interphase, CEP57 was expressed as shows a number of spots and dots proximal to the nucleus (Figures 4.26A and B). During prophase, CEP57 was expressed as a cluster of spots proximal to the condensed chromosome (Figure 4.26C). In prometaphase, CEP57 was expressed as shows a cluster of spots proximal to the mitotic chromosomes (Figure 4.26D).

CEP72 tagged with GFP tagged at the N-terminal was expressed in HEK293T cells. During interphase, CEP72 was expressed as groups of spots throughout the cytoplasm (Figures 4.27A and B).

CEP85 expression tagged with GFP at the C-terminal was expressed in HEK293T cells. CEP85 was expressed as two spots proximal to the nucleus in interphase (Figure 4.28A). During prophase, CEP85 was expressed as two spots proximal to the mitotic chromosomes (Figure 4.28B). In prometaphase, CEP85 was expressed as two clumps localised to opposite poles of the cell proximal to the mitotic chromosomes (Figure 4.28C).

PCM1 tagged with GFP at the N-terminal was expressed in HEK293T cells. During interphase, PCM1was expressed as one or two spots proximal to the nucleus (Figure 4.29A and B). In prophase, PCM1 was expressed as a spot proximal to the condensed chromosomes (Figure 4.29C). In prometaphase, PCM1 was expressed at a large spot proximal to the mitotic chromosome in the cytoplasm. PCM1 expression provides a good example of the life cycle of the centrosome during cell division.

Known cell cycle genes tested



Known cell cycle genes tested

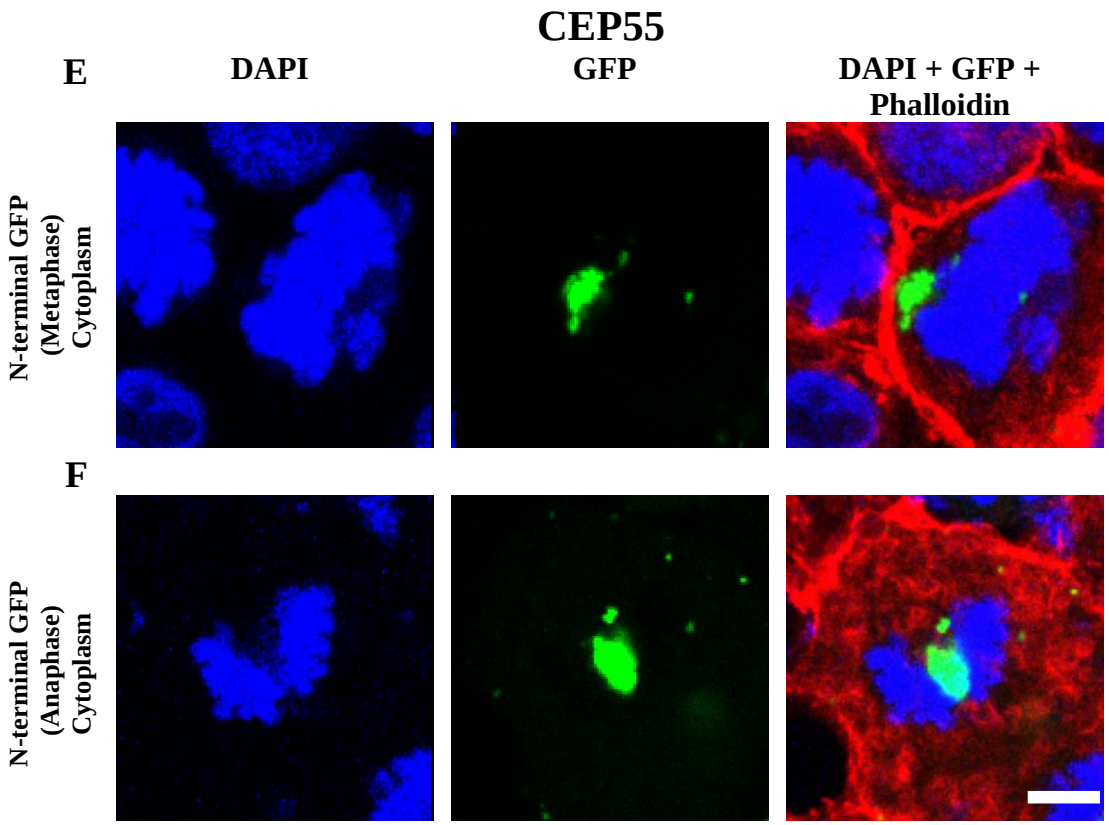


Figure 4.25: Subcellular localisation of CEP55 in HEK293T cells. Expression of CEP55 was proximal to the nucleus and mitotic chromosomes during **A)** interphase, **B)** prophase, **C), D)** prometaphase, **E)** metaphase, and **F)** anaphase. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

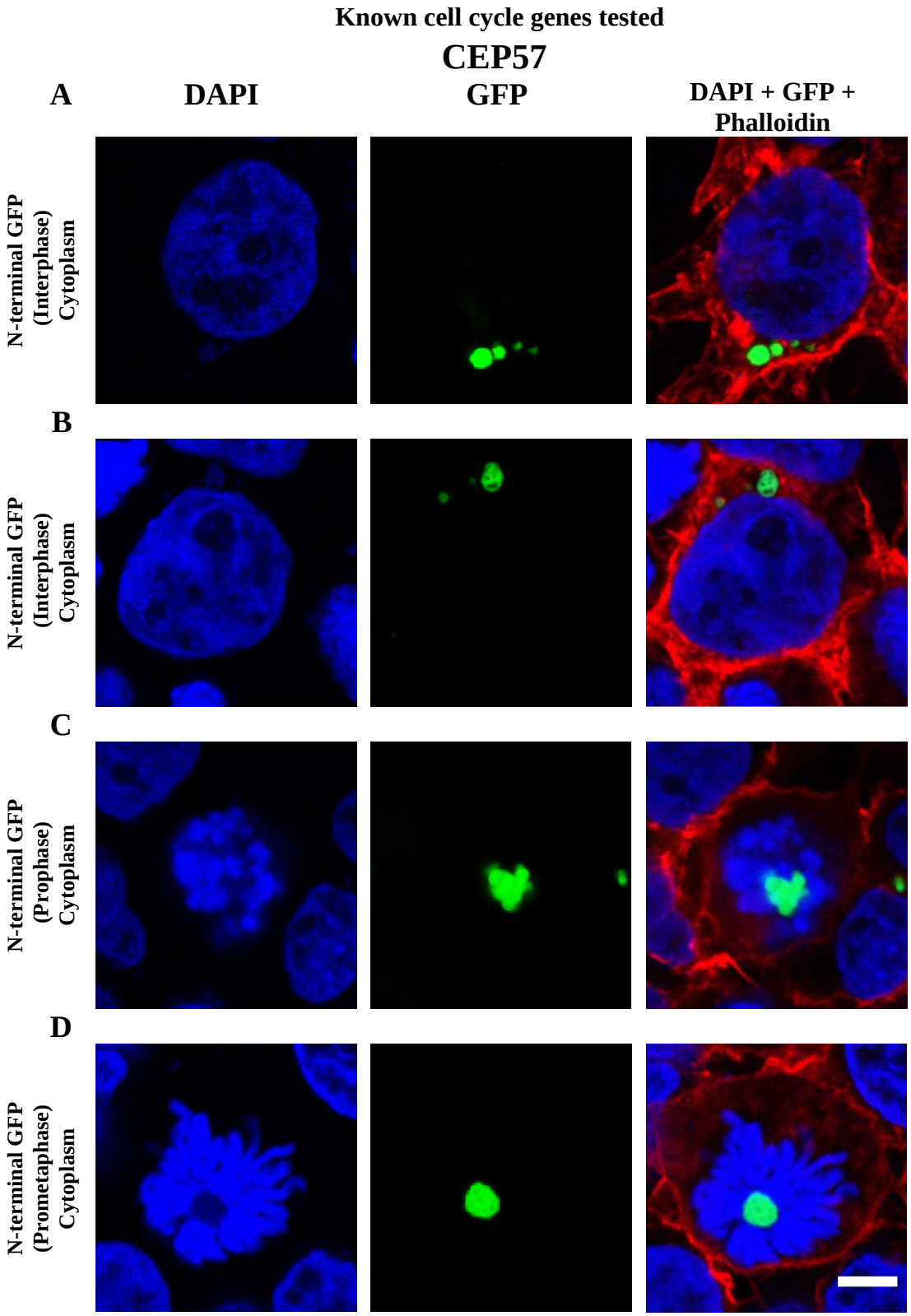


Figure 4.26: Subcellular localisation of CEP57 in HEK293T cells. Expression of CEP57 tagged with GFP at the N-terminal during interphase **(A** and **B)**, prophase **(C)**, and prometaphase (**D**). Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

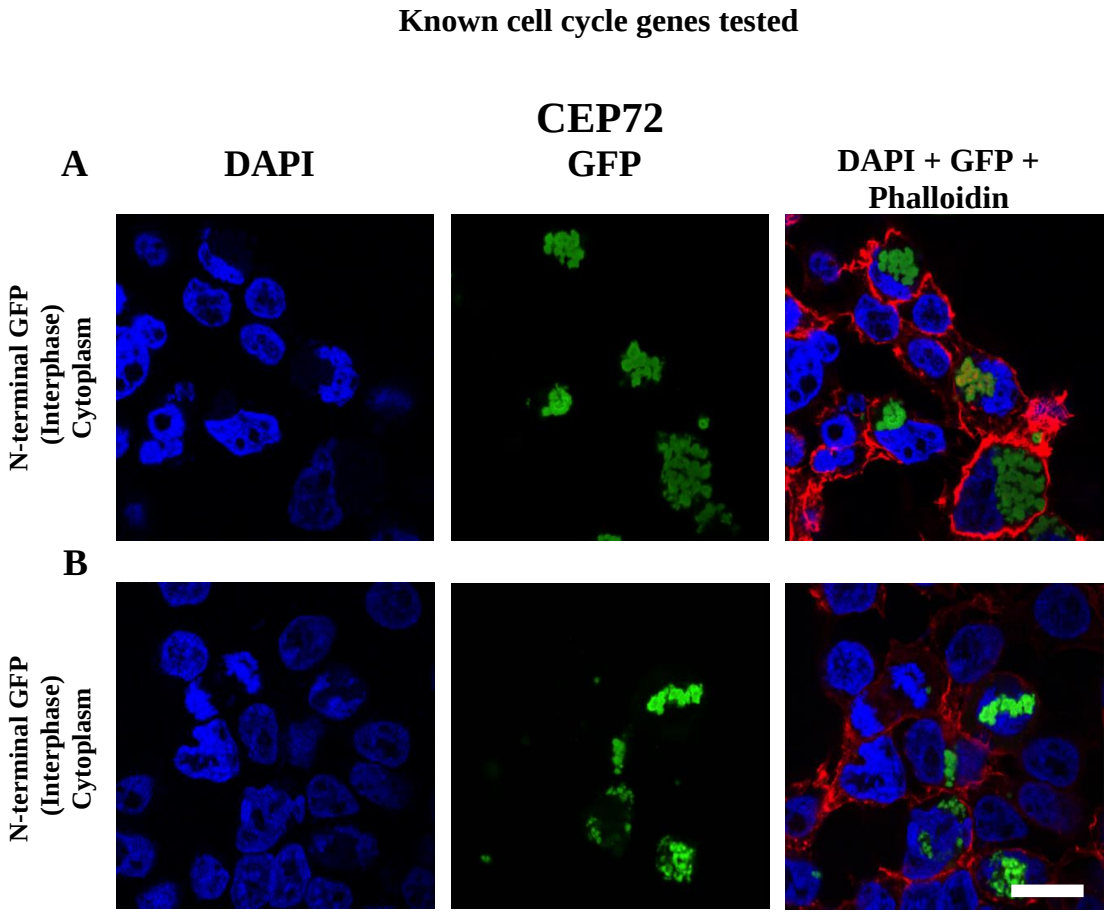


Figure 4.27 Subcellular localisation of CEP72 in HEK293T cells. A), **B)** Expression of CEP72 tagged with GFP at the N-terminal was localised to the cytoplasm. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 10 µm.

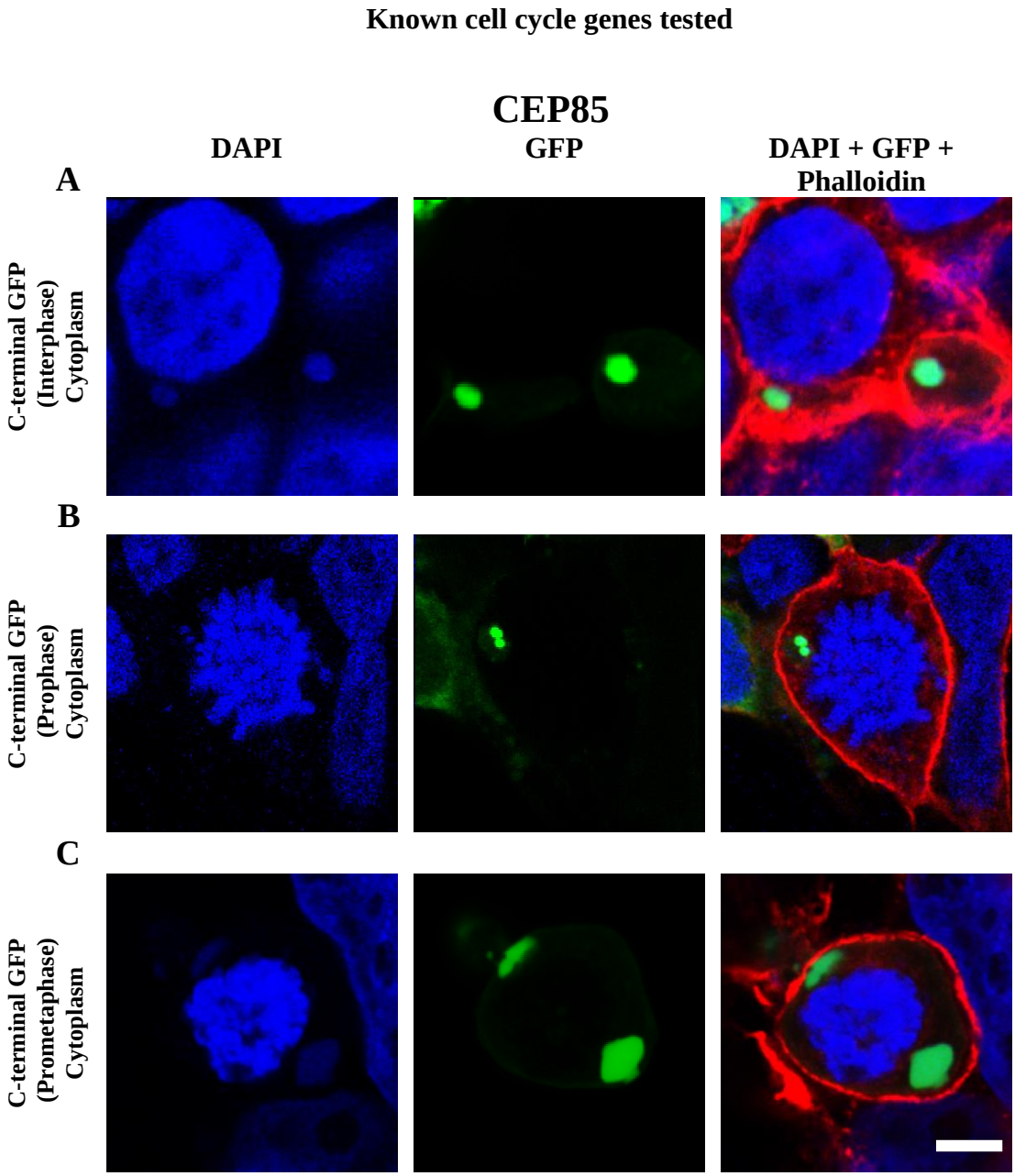


Figure 4.28 Subcellular localisation of CEP85 in HEK293T cells. A) Expression of CEP85 tagged with GFP at the C-terminal during interphase, **B),** and **C)** show the localisation of CEP85 expression during prophase and prometaphase, respectively. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

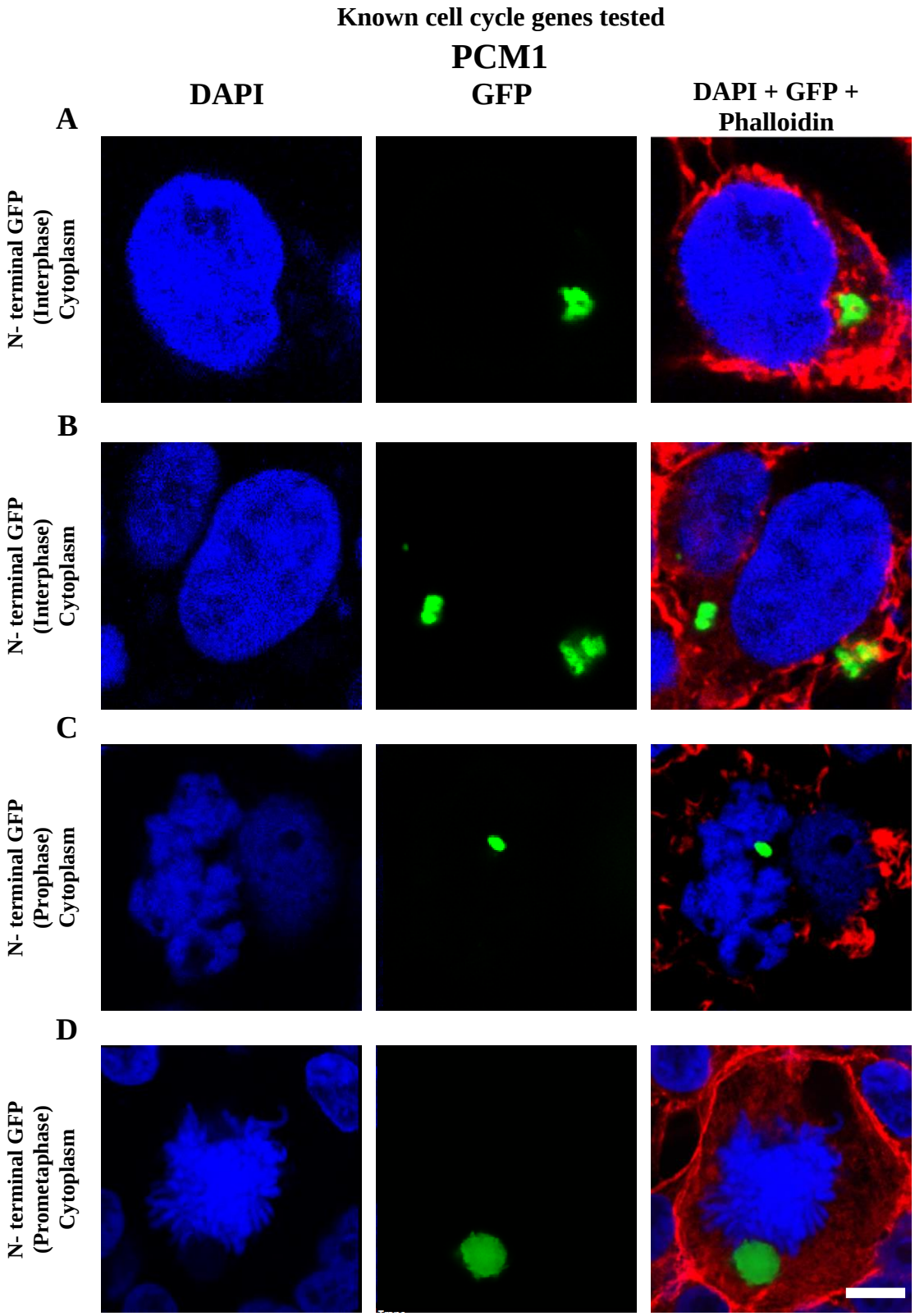


Figure 4.29 Subcellular localisation of PCM1. Expression of PCM1 tagged with GFP at the N-terminal in HEK293T cells during **A)**, **B)** interphase **C)** prophase, and **D)** prometaphase. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 4 µm.

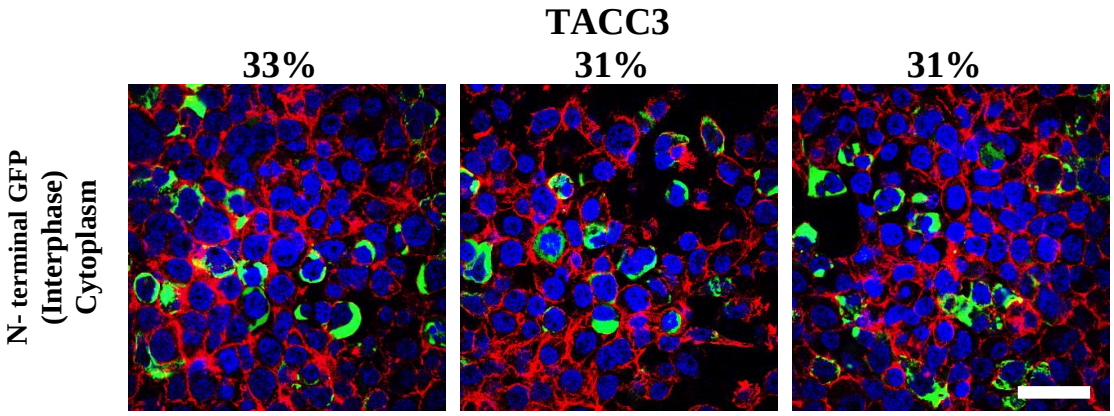


Figure 4.30 Subcellular localisation of TACC3 tagged with GFP at the N-terminal in HEK293T cells was used to determine the transfection efficiency. Cell images were captured by confocal microscopy (Nikon EC-1) 1000X and are 1.6 µm optical sections. Scale bar 20 µm.

The transfection efficiency was calculated using TACC3 tagged with GFP at the N-terminal. The number of cells expressing GFP in a field of view was composed of the total number of cell in the field of view. Transfection efficiency was found to be 30% (Figure 4.30).

Discussion

Subcellular protein localisation using GFP-fusion proteins in HEK293T cells was performed to provide evidence that the proteins of interest are involved in specific biological events during the cell cycle e.g. chromosome segregation conducted by the centrosomes during M-phase. Out of a total of 706 clustered genes that appear to be associated with transcriptomic activities in S-, G_2-, and M-phase, 69 genes were of unknown function (see Chapter 2). Of these genes of unknown function, 42 genes were selected according to the quantitation of their expression profile (see Chapter 2 methods). Screening by RNAi (see Chapter 3) resulted in identification of 19 of these genes with evidence for a role in proliferation (Chapter 3). Of these 19 genes, we were unable to obtain 7 clones (ARHGAP11B, CCDC77, FAM72B, GPN3, THAP10, TMEM194A, and ZNF367), whilst the other 12 clones were obtained and subcloned into expression vectors to study the localisation of the proteins in the cells using GFP (ATL2, C17orf53, ARHGAP11A, FIGNL1, FAM111A, RIBC2, KIAA1549L, RPL39L, REEP4, GSTCD, UBR7, and DEPDC1B).

In addition, I tested another 8 genes from the S, G_2/M-phase cell cycle list (Chapter 2). Of these genes, 3 are known cell cycle genes (i.e. ANLN, KIAA0101, and KNSTRN) and 5 are genes of unknown function (i.e. DEPDC1, FAM111B, KPNA2, RACGAP1P and TCF19). KIAA0101 and FAM111B had been previously screened in the RTCA experiment (Chapter 3) but their knockdown had not affected cell proliferation.

During the course of this project, reports were published that 10 of these proteins are associated with the cell cycle. It has been shown that ARHGAP11A/p53 complexes arrest the cell cycle and induce apoptosis [303], FIGNL1 is involved in DNA homologous recombination repair [308], REEP4 has a role in ER clearance and nuclear envelope reassembly [296], ANLN is involved in abscission (the final part of cytokinesis) [312], DEPDC1 regulates mitotic progression [313], KIAA0101 is involved in DNA repair [314], KNSTRN conducts the metaphase to anaphase transition [315], KPNA2 promotes cell proliferation [315] and TCF19 is necessary for cell proliferation [316]. The fact that ten proteins from the list of proteins of unknown function have been characterised as cell cycle proteins during this period represents very good evidence for the validity of the cell cycle list (Chapter 2).

In order to localise the expression of proteins within cells, the use of GFP-tagged proteins is a widely used technique [317]. The main limitation of this techniques is that tagging of

proteins with GFP may alter the localisation of the protein [317]. To minimise the potential for localisation artefacts, each protein can be tagged with GFP at the N- or C- terminal [317]. In the current study, the localisation of the GFP tag affected the protein localisation of FAM111A, KIAA1549L, and DEPDC1. In order to confirm the localisation of these proteins, immunofluorescence could be used to provide additional evidence [317].

It is interesting to note that for the proteins where localisation was not seen in M-phase cells (ARHGAP11A, KIAA1549L, RPL39L, GSTCD, UBR7, DEPDC1B, ANLN, DEPDC1, KIAA0101, KNSTRN, RACGAP1P, and TCF19), it is possible that the expression of these proteins may arrest the cells at interphase or result in apoptosis. This would account for the failure to observe cells expressing GFP-fusion proteins in mitosis. Further analysis e.g. flow cytometry and SYTOX® Blue staining (see Chapter 3) could be used to determine the cell cycle population and viability.

In general, M-phase duration is 3-5% of the total time of the cell cycle [318]. As the doubling time of HEK293T is between 16 to 24 h, the M-phase duration would be on average less than 1 hour. It might be that localisation in M-phase was not detected due to there being too few cells expressing GFP during this phase, the result of a lower transfection rate. Therefore, using cell cycle synchronisation techniques to generate a synchronous major population of cells might be useful to identify localisation of these proteins in M-phase cells.

As a member of the dynamin-superfamily, it has been reported that ATL2 is localised to the ER and involved in protein trafficking between the ER and Golgi apparatus [304]. In addition, it was reported that the knock down of ATL2 resulted in morphological changes of the Golgi apparatus and ER [304]. This suggests that ATL2 may be heavily involved in dynamin regulation that delivers the proteins from the ER to the Golgi. The localisation of ATL2 in Rismanchi *et al.*, 2008, showed numerous green spots (protein localisation fused with GFP at the N- or C-terminal) throughout the cytoplasm in HeLa cells which is different to the localisation pattern identified in this thesis where only several small patches of expression were observed in the cytoplasm. The C-terminal GFP fusion was not expressed in my study. Further experiments would use GFP fusion at the C-terminal in localisation experiments to test whether this agrees with the N-terminal protein localisation or the data of Rismanchi *et al.*, 2008. It is possible that the protein trafficking mechanism varies in HEK293T and HeLa cells, resulting in different patterns of localisation.

To further determine protein localisation, further analysis is required using co-localisation with known centrosomal proteins (e.g. γ-tubulin) or immune staining of specific antigens to identify particular cellular organelles.

C17orf53 appears to be localised throughout the cells in interphase, mitosis, and cytokinesis. This result was achieved by tagging with GFP tagged at the N- or C-terminal (Figure 4.6). It is interesting that C17orf53 is either localised to the cytoplasm as one or two spots proximal to the nucleus or as numerous spots throughout the cells. To further validate the localisation of C17orf53, co-localisation with known centrosomal proteins (e.g. γ-tubulin to determine if the spots of expression proximal to nucleus co-localises with centrosome) could be performed. The two distinct localisation patterns observed during interphase with both fusion proteins might be due to a difference in localisation at different time points during the cell cycle, this could be investigated by tracking cells expressing fusion proteins over a 24 h cycle.

It was reported that the function of FIGNL1 has been characterised during the course of this thesis, it has been shown to be involved in homologous recombination repair during S phase [308]. In addition, previous studies demonstrated that FIGNL1 is localised to both the nucleus and cytoplasm, and potentially plays a role in cell proliferation and differentiation [319]. However, FIGNL1 in this experiment appears to be localised to the cytoplasm as one or two spots proximal to the nucleus in interphase and mitosis (Figure 4.8). The C-terminal GFP fusion was not expressed in this study. To further study the localisation of FIGNL1, additional localisation of proteins fused with GFP at the C-terminal is required.

FAM111A fused with GFP at the C-terminal was localised to discrete patches of expression proximal to and within the nucleus (Figure 4.9C). However, when FAM111A was fused with GFP at the N-terminal, it was expressed in discrete patches proximal to the nucleus and at lower levels throughout the cytoplasm (Figures 4.9A and B). It is most likely that this difference in expression is due to the N-terminal fusion protein being expressed at a higher level. Tagging of GFP at different position of proteins might also influence the protein trafficking system in HEK293T cells e.g. the discrete nuclear patches of expression observed only with the C-terminal fusion protein. Further experiments might use immunofluorescence if suitable antibodies were available. There is little data published for FAM111A but a previous study suggested that FAM111A plays an important role in viral host range restriction i.e. depletion of FAM111A may increase SV40 viral gene expression [320]. It might be therefore follow that FAM111A has a function in inhibiting the cell cycle in

response to viral infection. However, this is not what we observed in the knockdown experiments in Chapter 3 where a decrease in FAM111A expression resulted in a decrease in cell proliferation. Further experiments might investigate the role of FAM111A on the cell cycle whilst infected with a virus.

FAM111B appears to be localised to the position of the centrosome in the current study (i.e. consistent pattern of FAM111B localisation as a discrete patch proximal to the nucleus when fused with GFP at the N- or C-terminal) (Figure 4.19). Little is known of the biological role of FAM111B in the cell cycle or other mechanisms. To confirm the centrosomal localisation of FAM111B, additional analysis is required e.g. co-localisation with a known centrosomal protein such as γ-tubulin.

KIAA1549L fused with GFP at the N-terminal appears to be localised to the position of the centrosome during interphase (Figure 4.11A). However KIAA1549L fused with GFP at the C-terminal appears to be localised throughout the nucleus and cytoplasm (Figure 4.11B). To determine whether the localisation with the N-terminal GFP fusion protein is co-localised with the centrosome, it is necessary to perform co-localisations with a known centrosomal protein e.g. γ-tubulin. To determine which of the conflicting patterns of localisation is correct, analysis using immunofluorescence could be performed if a suitable antibody was available.

Recent studies suggest that REEP4 is involved in the regulation of microtubule binding with chromosomes during mitosis [296]. It was reported that REEP4 also conducts the NEBD by removing the mitotic ER from metaphase chromosomes and therefore playing an important role in the cell cycle [296]. REEP4 was characterised during the current study, and has been reported to be localised to the ER [296]. REEP4 appears to be localised to the cytoplasm in this thesis and therefore shares a similar location with the published evidence i.e. localised to the cytoplasm.

As a member of the karyopherin family, it has been reported that KPNA2 functions as a nucleocytoplasmic transporter to deliver p65 (transcriptional activator) to the nucleus [321]. In addition, KPNA2 is reported to be overexpressed in epithelial ovarian carcinoma (ECO) promoting the G_1/S cell cycle transition [322]. The knockdown of KPNA2 effectively suppressed the proliferation of ECO [322]. In this thesis, KPNA2 appears to be localised to the cytoplasm proximal to the nucleus as one or two bright spots when tagged with GFP at the N-terminal. However, KPNA2 localisation was not identified in M-phase and the C-

terminal GFP fusion was not expressed in this study. To further determine its localisation in the cells, expression of c-terminal GFP fusion proteins should be performed. Co-localisation with known cell cycle-associated proteins e.g. CEP55, and immunofluorescence studies could also be performed.

KIAA0101 appears to be localised to the chromosomes appearing as dashed line patterns throughout the nucleus. It has been reported that KIAA0101 is a proliferating cell nuclear antigen (PCNA) binding protein [314]. PCNA plays a role in DNA replication and damage repair during S-phase. Since DNA repair is conducted by a protein complex composed of PCNA and KIAA0101, the overexpression of KIAA0101 fused with GFP at the N-terminal could result in a thick bright green dashed line in the nucleus (Figure 4.20A). Both KIAA0101 fused with GFP at the N- and C-terminal are localised to the nucleus. However, the localisation pattern of KIAA0101 fused with GFP at the C-terminal is different compared to the expression of KIAA0101 tagged with GFP at the N-terminal. The expression of KIAA0101 fused with GFP at the C-terminal appears to be localised as numerous dots throughout the nucleus (Figure 4.20B).

ARHGAP11A appeared to be localised to the nucleus during interphase and to the chromosomes during metaphase and cytokinesis. The C-terminal GFP fusion was not expressed in this study and further experiments would examine this localisation. ARHGAP11A is reported to be involved in cell cycle arrest by binding with p53 [303]. Combination with ARHGAP11A stabilises the tetrameric conformation of p53, enhancing the binding with DNA and eventually resulting in cell cycle arrest [303].

As a member of the APC/C, ANLN plays an important role in cell cycle progression [312]. During cytokinesis, ANLN has been reported to conduct the separation of daughter cells by assembling an actin and myosin contractile ring [312]. In addition, ANLN has been found to be primarily localised to the nucleus in interphase and appears to be localised to the cell cortex upon nuclear envelope breakdown [312]. The cell cortex is a specialised layer of cytoplasmic protein assemblies on the inner face of the plasma membrane at the cell periphery. ANLN has also been found to be localised to the cleavage furrow during cytokinesis [312]. This is not what was observed by expressing the N-terminal GFP fusion in HEK293T cells where expression was observed as a large mass in the cytoplasm (Figure 4.17). The C-terminal GFP fusion was not expressed in these studies and further experiments would examine this localisation.

DEPDC1B is localised to the nucleus and cytoplasm (Figure 4.16). In contrast to the localisation pattern of DEPDC1B, DEPDC1 appears to be localised to the cytoplasm and the nucleoli but not the nucleus, which is consistent with the published literature [313]. DEPDC1B is involved in the de-adhesion events of cells in mitosis [307]. The knockdown of DEPDC1 resulted in mitotic arrest suggesting it is heavily involved in cell cycle regulation [313]. It appears to that DEPDC1 and DEPDC1B are both heavily involved in M-phase activities.

KNSTRN appears to be localised to the cytoplasm as grouped spots in interphase (Figure 4.21). KNSTRN is reported to be an essential component of the mitotic spindle which is heavily involved in chromosome segregation [315]. It is also involved in the transition between metaphase and anaphase, which is important in the preparation for chromosome segregation [315]. However, the localisation of KNSTRN was not identified in mitotic cells in this experiment. It is possible that in rapid dividing cells e.g. HEK293T the time window of M-phase shifts rapidly. To identify the localisation pattern in mitotic cells, it would be useful to localise KNSTRN in cells after a synchronisation treatment.

TCF19 appears to be localised to the nucleus as a number of green spots (Figure 4.24). Recent experimental approaches suggest that the knockdown of TCF19 may decrease cell proliferation and induce apoptosis in an insulinoma cell line [316]. In addition, it has been reported that the knocking down of TCF19 results in the down regulation of cyclin A, B, and E suggesting it is heavily involved in cell cycle regulation [316].

RPL39L appears to be localised to the nucleus when it is tagged with GFP at the N-terminal (Figure 4.12). However, it has been reported that the sequence of RPL39L is known to be highly homologous to RPL39 (ribosomal protein L39) suggesting that RPL39L may also be localised to the ribosome in the cytoplasm [323]. It is possible that GFP tagging at the N-terminal affects the protein trafficking proteins in HEK293T cells. To confirm the protein localisation with the N-terminal fused GFP, further studies would perform localisation of protein with GFP tagged at the C-terminal.

In summary, main conclusion for this chapter is that FAM111B and KIAA1549L were found to be localised to the position of the centrosome. To provide additional evidence, further studies would test co-localisation with γ-tubulin to confirm this. However, the role of these genes in centrosome regulation remains unknown. In order to begin to understand the role of proteins of unknown function in the $S/G_2/M$ gene list, it was thought that a large-scale

pathway diagram of protein-protein interaction in the centrosome regulatory network would be of great benefit. The construction of this diagram is described in Chapter 5.

Chapter 5.

Construction of a large-scale molecular model of centrosome and its duplication during cell division

Introduction

Protein localisation (Chapter 4) analyses produced a list of proteins, which may be associated with the centrosome e.g. FAM111B. In order to depict the putative position of these poorly characterised proteins in the centrosome, a large-scale pathway map was constructed.

Centrosomes are known to play an important role in cell division. During prophase, the centrosomes migrate to the opposite poles of the cell. The nuclear membrane then breaks down in preparation for mitotic spindle assembly, which connect the centrosome and chromosome. Upon microtubule contraction condensed sister chromatids segregate into two daughter cells [324].

Under the regulation of CDK2, centrosomes only duplicate once per cell cycle during S-phase giving each daughter cell one centrosome [325]. Each centrosome normally consists of two centrioles, which are surrounded by PCM. Centrioles are composed of triplet microtubules arranged as barrel-shaped structures. Paired parental centrioles that carry distal appendages can be distinguished by staining the marker proteins, such as centrosomal protein 164kDa (CEP164) [326,327]. The centrosome regulation is the primary microtubule-organizing centre and governs protein-protein interactions that occur in interphase and mitosis [328].

A biological process can be described by text and sentences in papers e.g. the cyclic adenosine monophosphate (cAMP) signalling pathway [329]. However, it is difficult to describe a large-scale regulatory system e.g. the cell cycle control system [175], and immune response [174] that includes many protein acting as a complex network.

In the past 30 years, huge amounts of data have been generated in the field of molecular biology, including information on individual proteins, and their interactions within a cell. These protein characteristics and interactions reported in the literature have been curated

centrally, such as KEGG [330] and REACTOME [188] or by the community, including WikiPathways [331].

A pathway map could simply be considered as a network composed of many workflows of protein-protein interactions, linking upstream and downstream interactions [177]. In a pathway map, the proteins are considered as components and the protein interactions are depicted as arrows between components. For further information of pathway map please see pathway modelling at page 37 in Chapter 1.

Over the last decade, several approaches for representing pathway maps have been proposed. The SBGN is a graphical representation language for visualising the biological processes as a pathway map [179]. The mEPN is an alternative graphical language based on a pathway notation system first proposed by Moodie *et al.*, [176] in 2006 and first published in its current form by Raza *et al.*, 2008 [174].

The SBGN and mEPN schemes are similar in concepts; components, processes, and cellular compartments used to construct pathway nodes in both schemes. SBGN based pathway mapping is developed by various members of the community, such as the REACTOME, and the mEPN is developed by a team of scientists at the Division of Pathway Medicine and the Roslin Institute in the University of Edinburgh.

The aim of this chapter is to construct a detailed centrosome-associated protein network by studying existing models and published literatures. In this study, I have constructed a large-scale regulatory of the centrosome life cycle, which is used to improve current understanding of centrosome biology.

Materials and methods

Data mining, information curation, and organisation

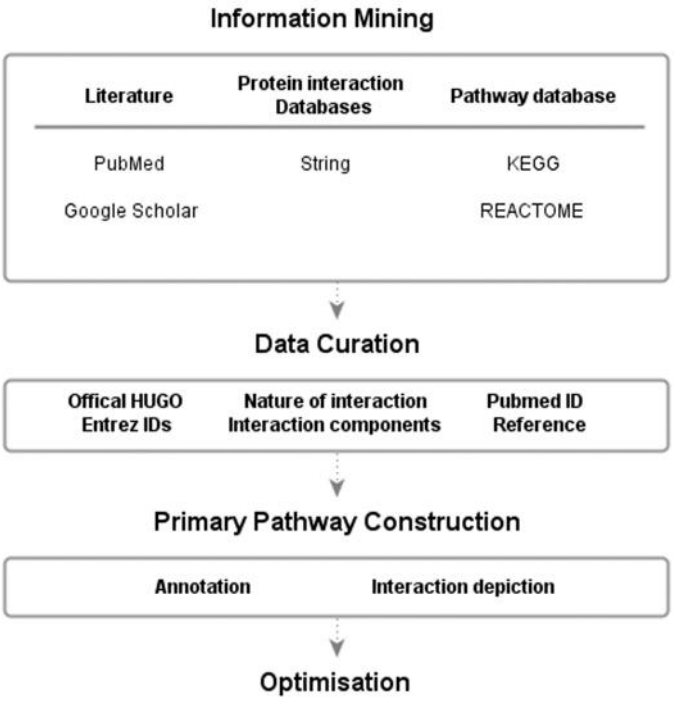


Figure 5.1: Flow chart of pathway mapping. The flow chart summarizing the assembly of information across three main stages from concept to diagrams. The tools and methods used at each stage are listed in the grey boxes.

A scaffold pathway map was composed of well-characterised centrosome-associated proteins. PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and Google scholar (http://scholar.google.co.uk/) were used to perform searches for literature on centrosome proteins involved in interphase and during mitosis (Figure 5.1). Data mining on known protein-protein interactions between centrosome proteins was carried out using online databases, including String (http://string-db.org/) [332]. In addition, online pathway map databases, such as KEGG and REACTOME were used to provide the pathway maps in which proteins with additional unknown functions may be involved.

All components have been depicted based on standard human gene nomenclature (HGNC) [333]. For each relevant interaction, the following details were recorded: the PubMed IDs of the publication the reported the protein-protein interaction, the Enterz IDs of the proteins, and the nature of the interaction (see Appendix 1) [177].

Pathway construction

The entire protein-protein interaction network in this pathway was built as directional networks and finished in 6 months. The rules and details are provided in www.mepnpathway.org and current mEPN scheme [177].

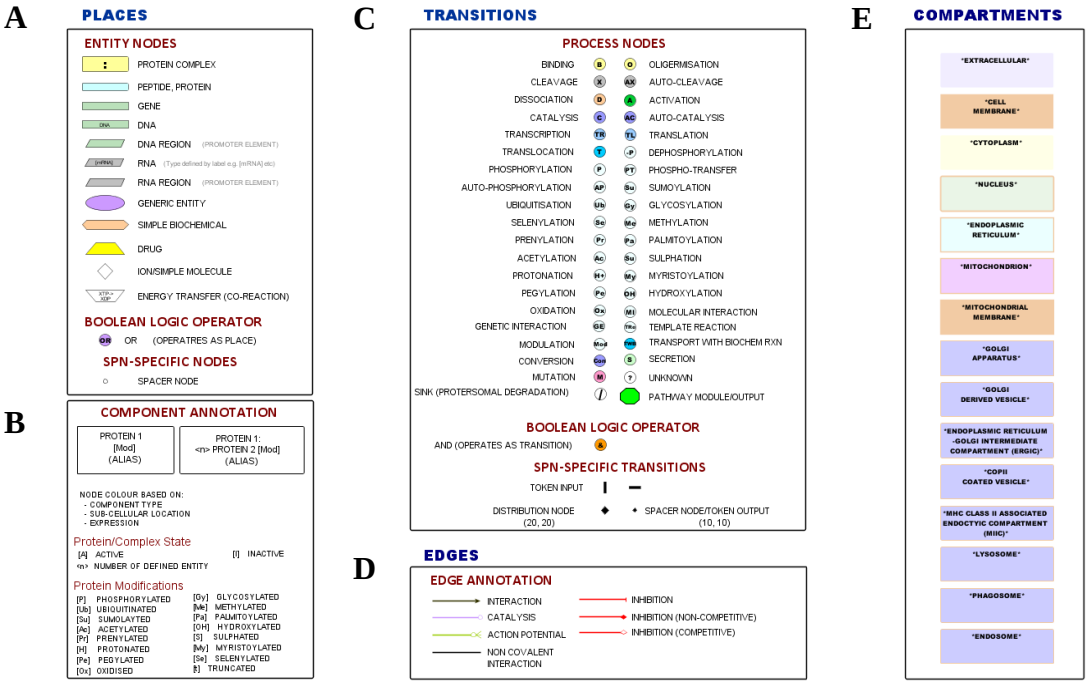


Figure 5.2: List of the glyphs used by the modified Edinburgh Pathway Notation (mEPN). Different shapes and identifiers are adopted to distinguish between each element of the notation scheme. The notation scheme consists of the following categories of **A)** nodes representing 'place' for subcellular location, **B)** the state of components, **C)** 'Transitions' for protein-protein interactions, and 'Boolean logic operators' for the relationship between multiple inputs into a biological process, **D)** 'Edges' for the biological regulation in the cells. **E)** 'Compartments' for the cell and locations. For full scale images please see attached PDF file in CD.

The entire pathway of centrosome-regulated proteins can be considered as a network of interactions between components. To visualise knowledge from literature systematically, the mEPN scheme is explained here. Graphs can be classified into four categories. First, entity nodes are used to represent any physical entity that is involved in the biological interactions, such as a protein, protein complex, nucleic acid, and other molecules (Figure 5.2A). To further display the state of proteins, the abbreviation of descriptions within the brackets can be used e.g. [A] means activate, [P] means phosphorylated (Figure 5.2B).

Transition nodes, such as 'B' means the binding between proteins, 'P' means the phosphorylation can represent interactions between components in the pathway (Figure

5.2C). Direction of biological interactions are depicted with arrows of various colours and shapes (Figure 5.2D). To define the position where these biological processes happens, compartments are used to define the location in cells, such as cytoplasm and nucleus (Figure 5.2E).

Pathway maps were constructed and optimised using graph editor program yEd (yFiles software, Tubingen, Germany). Information of components and their associated evidence of interactions between proteins, PubMed IDs are stored on appropriate edges and nodes. Pathways saved as '.graphml' files. The optimisation of the pathway map also involved redirecting the edges that connect proteins interactions such as, catalysis, and inhibition.

Design of the centrosome regulation pathway in longitudinal and cross sectional displays

The aim of pathway mapping is to systematically incorporate and visualise information from literature curation. However, the number of components in a pathway determines the information accessibility i.e. understanding the regulatory logic of a single interaction is easier than the entire regulatory network of the centrosome life cycle which is composed of more than 100 proteins. To help users effectively access the pathway map, a flow chart that represents the centrosome regulation at each stage by longitudinal and cross sectional images is used here.

Pathway overlay of the transcriptional analysis of fibroblast before and after cell cycle synchronisation

To understand the regulatory logic of centrosome-associated genes, overlapping genes between the map and array dataset (Chapter 2) in this thesis are labelled with different colours according to their expression periodicities where they up- or down-regulated in the cell cycle.

Results

Construction of a framework diagram of centrosome regulation pathways

Of those proteins associated with centrosomes, 117 were selected as having the most significant involvement in centrosome regulation based on literature mining. Other associated proteins, which are not necessarily involved in the regulation of centriole duplication were not included in this pathway map.

The pathway map of centrosome regulation includes 667 nodes of which 270 represent proteins, 203 complexes, and 192 make up other components. A total of 93 interactions are described in the pathway map, of these interactions 37 are various activation state modulations (20 phosphorylations, 3 dephosphorylations, 6 cleavages, 2 translocations and 6 activations by processes not defined).

There are 6 inhibition reactions, 5 of these are inhibition of cleavage while the other inhibits microtubule nucleation. A total of 2 translocation events occur during disjunction of paired parental centrosomes. In many circumstances the same paper may describe multiple interactions, for example the recruitment of CEP55 to the midbody during cytokinesis was inhibited by PLK1 [43].

To describe centrosome duplication at different stages in the cell cycle, different terms are used (see below). The developing of procentrioles happened on both of mother and daughter centriole during interphase (Figure 5.6 and Appendix III). It is important to note that during M-phase, both mother and matured daughter centriole are separated at opposite positions with its own procentrioles wrapping with PCM (Figure 5.6 and Appendix III). To specifically point out which stage where there are, 'parental centrosome' is used to define the M-phase centrosomes.

• G_1-early	Mother centrosome:	A completely developed centriole with appendages wrapped with PCM.
• G_1-late & S	Procentriole:	A daughter centriole undergoing the duplication during S-phase.
• S	Interphase centrosome:	A matured mother centriole wrapped with PCM and a procentriole.
• S & G_2	Daughter centriole:	A completely developed centriole by the end of S-phase.
• G_2	Daughter centrosome:	A daughter centriole just dissociated from its mother centrosome 'wrapping with PCM'.
• M	Parental centrosome:	Once the cell enters M-phase, two matured centrosomes are ready to function as mother (parental) centrosomes after moving to two duplicated cells.

For important elements of the centriole structure, there are 2 proteins located in procentrioles, 3 in proximal end of the centriole, 8 in distal end, 10 in 9 symmetrical microtubules, and 9 in appendages (Table 5.1). The PCM is consisted of various areas including inner layers (4 proteins), outer layers (6 proteins), basement of daughter centriole (4 proteins), basement of mitotic spindle assembly (11 proteins), and centriolar satellite (12 proteins) (Table 5.1). The linker fibre includes 5 proteins (Table 5.1). The cohesion complex includes 9 proteins (Table 5.1). The other centrosomal proteins are located in the cytoplasm (26 proteins) and midbody (9 proteins) (Table 5.1).

To understand the life cycle of centrosome during cell proliferation, figures that explain the structure of centrosome from the literature are shown (Figure 5.3). The following figures were the base for building the pathway map. For the longitudinal section of the diagrams, figure 5.3A shows the structure of the mother and the daughter centrioles according to the physical position of the proteins' location [324].

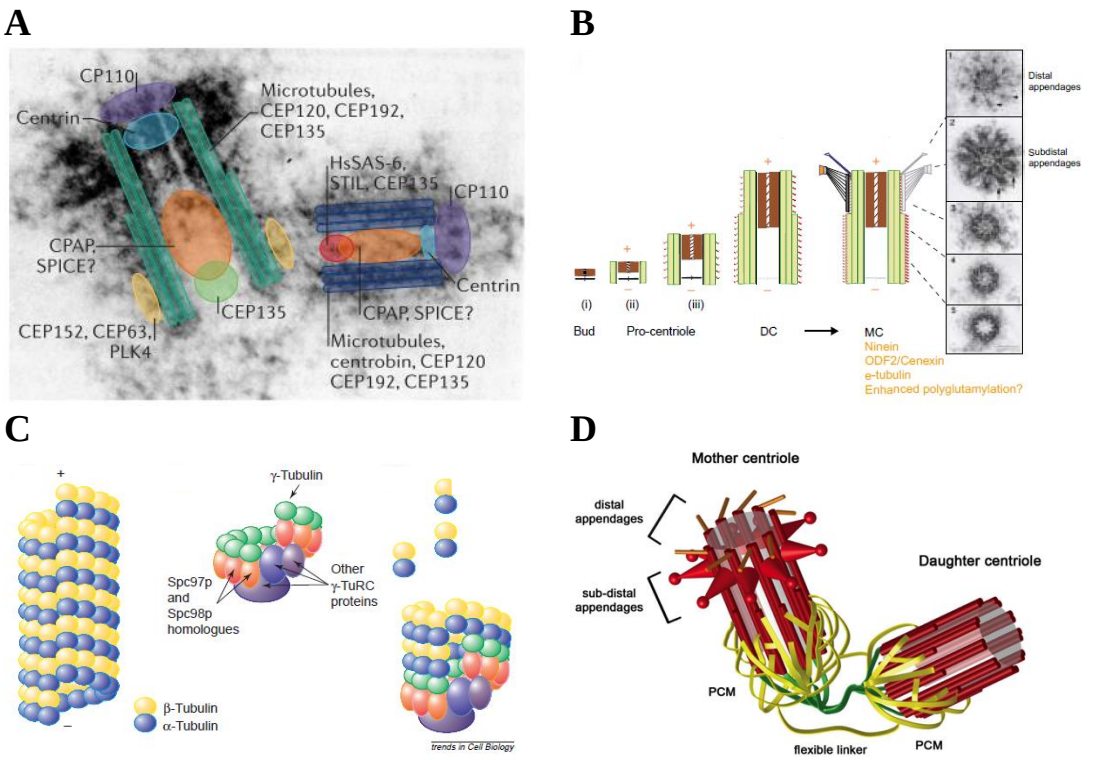


Figure 5.3: Examples of centrosome structure in 2D and 3D. A) The electro micrograph demonstrates an interphase centrosome including a mother and daughter centriole at longitudinal section. This figure is courtesy of Gonczy *et al.*, 2012 [324]. **B)** The development of a mother centrosome is shown in a cartoon flowchart that includes the electron micrograph of centriole at proximal and distal end. This figure is courtesy of Bornens *et al.*, 2002 [334]. **C)** The microtubule assembly shown in 3D. This figure is courtesy of Oakley *et al.*, 2000 [335]. **D)** A mother and daughter centriole showed in 3D at structure level. This figure is courtesy of Sillibourne *et al.*, 2010 [336].

Figure 5.3B demonstrates the development of a centriole from the assembly of the cartwheel structure to the recruitment of appendages [334]. This figure also includes the electron micrograph demonstrating the centriole structure at different levels of cross section, which was used to depict cartoon style diagrams. Figure 5.3C explains the development of microtubule assembly including α, β, and γ-tubulin as the basic elements in a 3D environment [335]. Figure 5.3D shows the entire structure of mother and daughter centrioles wrapping with PCM in a 3D environment to provide more detail of centrosome at structure level [336].

Contemporary view on centrosomes structure and function

The centrosome is known to be composed of multiple functional complexes consisting of hundreds of proteins [337,338]. Of those proteins, 44 important proteins are believed to be essential in the structured core of a post-mitotic centrosome and are demonstrated in cross and longitudinal sections (Table 5.1 and Figure 5.4).

Centrosome structure

A fully developed centrosome consists of two centrioles surrounded by the PCM (Table 5.1, Figure 5.4A and D) [52]. More specifically, a mother centrosome is composed of a cartwheel structure [339], a 9-fold symmetrical structure of microtubule triplets (A, B, and C-tube) [340], a PCM from the centre to surrounding materials (Table 5.1).

Centriole	
Procentrioles	NPM1, ROCK2
Proximal end (Cartwheel structure)	CENPJ, CEP135, SASS6
Distal end	CETN1, ODF1, USP33, POC1B, CCNF, CCP110, POC1A, POC5
9 symmetrical microtubules	TUBGCP2, TUBGCP6, TUBE1, TUBA1B, TUBB3, TUBD1, TUBG1, TUBGCP3, TUBGCP4, TUBGCP5
Appendages	MAPRE2, MAPRE3, CEP170, ODF2, CEP164, CNTRL, MAPRE1, NIN, KIF24
PCM	
Inner layers	AKAP9, CDK5RAP2, CEP120, PCNT
Outer layers	CEP63, ASPM, CEP152, CEP57, CNN, NEDD1
Basement of daughter centriole	PLK2, CEP192, PLK4, STIL
Basement of mitotic spindle assembly	PAFAH1B1, NUMA1, HAUS5, HAUS1, HAUS2, HAUS3, HAUS4, HAUS6, HAUS7, HAUS8, KIF11
Centriolar satellite	DCTN1, DISC1, BBS4, HOOK3, OFD1, KIZ, CEP90, PARD6A, CEP290, CEP72, FOPNL, PCM1
Linker fibre	CROCC, CEP250, CEP68, LRRC45, CTNNB1
Cohesin complex	STAG3, PP2A, CC2D1A, RAD21, SGOL1, SMC1, SMC3, SPAG5, STAG2
Cytoplasm	CASS4, CDH1, PPP1CA, PPP1R2, DNAH2, TEKT1, NEK6, NEK9, NDEL1, SAV1, MAPK1, CCNE1, ANAPC, CDK1, FZR1, KATNB1, KPNB1, PLK1, TPX2, FBXW5, AURKA, NEK2, NEK7, PPP1CC, STK3, CDK2
Midbody	STAG3, PP2A, CC2D1A, RAD21, SGOL1, SMC1, SMC3, SPAG5, STAG2

Table 5.1: Proteins, which are involved in the centrosome biology in the pathway map.

Between the mother and daughter centrioles, the linker fibres provide dynamic connections by binding with their proximal ends (Table 5.1 and Figure 5.5) [341]. Compared to the daughter centriole, the mother centriole has additional distal and subdistal appendages (Table 5.1, Figure 5.4A and D) [52].

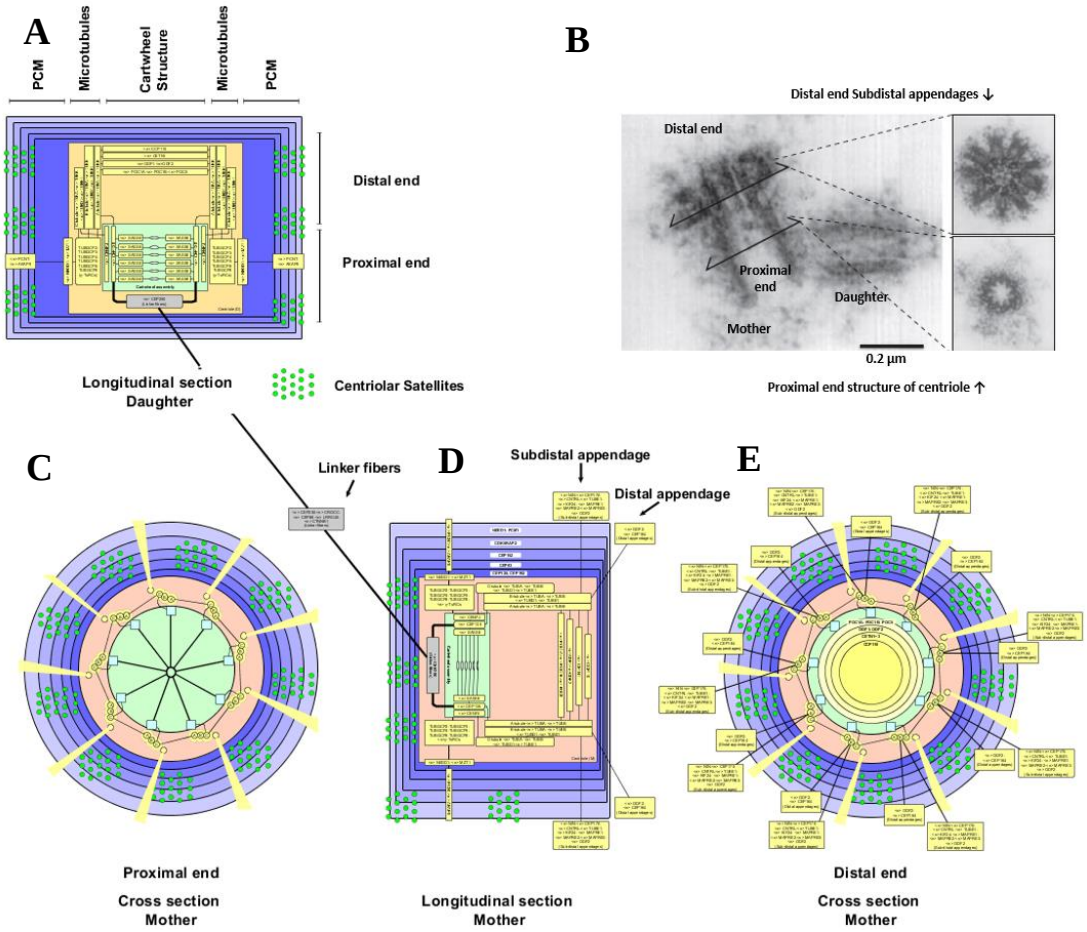


Figure 5.4: Centrosome structure. Cross and longitudinal sections of the interphase centrosome depicted according to the electron micrograph of centrosome. The structure includes matured mother, daughter centrioles, and a highly structured PCM. **A)** The longitudinal section of a daughter centrosome. **B)** Electron micrograph of the interphase centrosome. The top and bottom inset demonstrates the cross-section of subdistal appendages and proximal part of the centrosome. Image was generated by Paintrand, *et al.*, 1992 [342]. **C)** The cross section of a mother centrosome at proximal end. **D)** The longitudinal section of a mother centrosome. **E)** The cross section of a mother centrosome at distal end. For full scale images please see figure 5.5 and PDF file attached in CD.

Cartwheel structure and the proximal end of the centriole

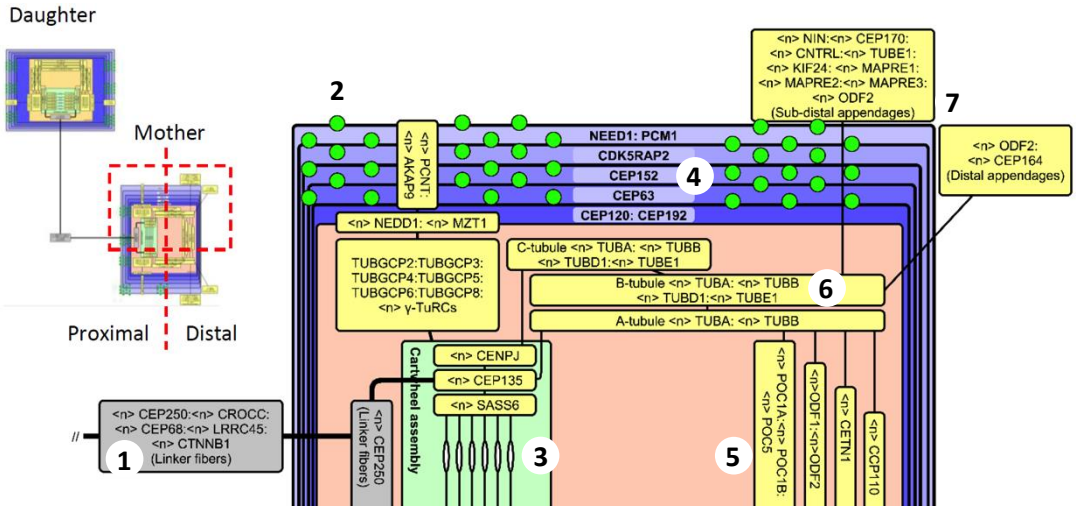
The cartwheel structure at proximal end of a centriole is composed of protein spindle assembly 6 (SASS6), CEP135, and centromere protein J (CENPJ) (Figures 5.5A3 and B2). Composed of nine spokes radiating from the centre hub, the cartwheel assembly at the proximal end of the centriole structure is entirely composed of protein SASS6 [340]. SASS6 is the central scaffolding component of the centrioles building a 9-fold symmetrical structure of microtubule triplets (A, B, and C-tube) [340]. In connection between SASS6 and microtubule triplet, CEP135 plays a role as the scaffolding structure during early stage of centriole cartwheel structure biogenesis [258]. CENPJ plays a structure role in maintaining the integrity of the cartwheel scaffold and the spindle morphology in connection with C-tube and CEP135 in cell proliferation [343].

Proteins located at the distal end of the centriole

At the distal end of the centriole structure, the protein complex includes centriolar coiled-coil protein of 110 kDa (CCP110), ubiquitin specific peptidase 33 (USP33), proteome of centriolar protein A (POC1A), proteome of centriolar protein B (POC1B), proteome of centriolar protein 5 (POC5), outer dense fibre of sperm tails 1 (ODF1), outer dense fibre of sperm tails 2 (ODF2), centrin-1 (CETN1) (Figures 5.5A5 and B3). CCP110 is a cell cycle-dependent CDK substrate regulating centrosome duplication [344]. Protein USP33 inhibits the ubiquitination of CCP110 progressing the centrosome duplication [345]. POC1A and POC1B were both involved in the early stages of centriole duplication stabilising the structure integrity of centriole and proper mitotic spindle formation after daughter centrosome fully developed [346]. POC5 is known to have an essential role in the centriole elongation at the distal half of centrioles [347].

Other distal end centriole proteins include ODF1 which regulates the length and distal structures of centrioles [348], and ODF2 is required for the formation of the distal and subdistal appendages on the mother centriole (Figures 5.4B, 5.5A5, and 5.5B3) [349]. Protein CETN1 is involved in the regulation of microtubule assembly, the coordination with the MTOC, and the duplication of centriole in mammalian cells (Figures 5.4B, 5.5A5, and 5.5B3) [350,351]. The phosphorylation of CETN during mitosis determines centriole disjunction between mother and daughter centrioles [352].

A



B

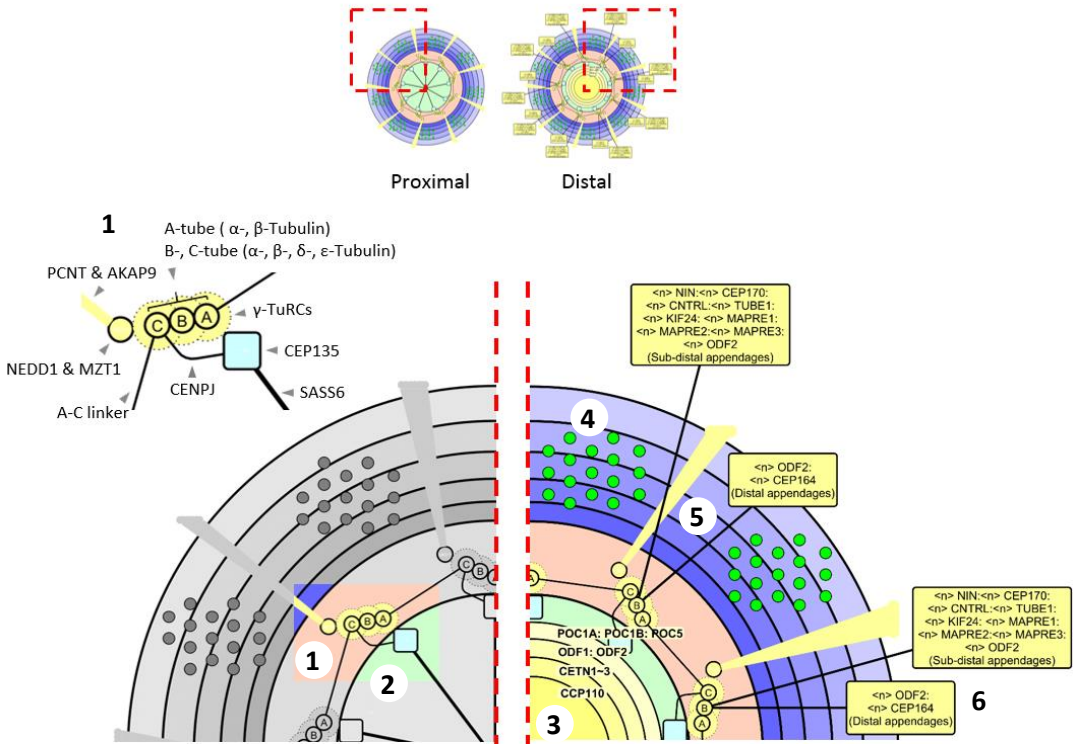


Figure 5.5: Longitudinal and cross section of a mother centrosome. A completely developed interphase centrosome includes **A1)** Linker fibres, **A2** and **B4)** Centriolar satellites, **A3** and **B2)** proximal end scaffolds, **A4** and **B5)** PCM, **A5** and **B3)** distal end scaffolds, **A6** and **B1)** 9 fold of symmetrical microtubule triplicate, **A7** and **B6)** distal/subdistal appendages. A matured mother centriole is paired with a daughter centriole via the connection of **A1)** linker fibre-associated proteins. The dashed black line separates the sections between the cartwheel structure **A3** and **B2)** and the microtubules **A6** and **B1)**.

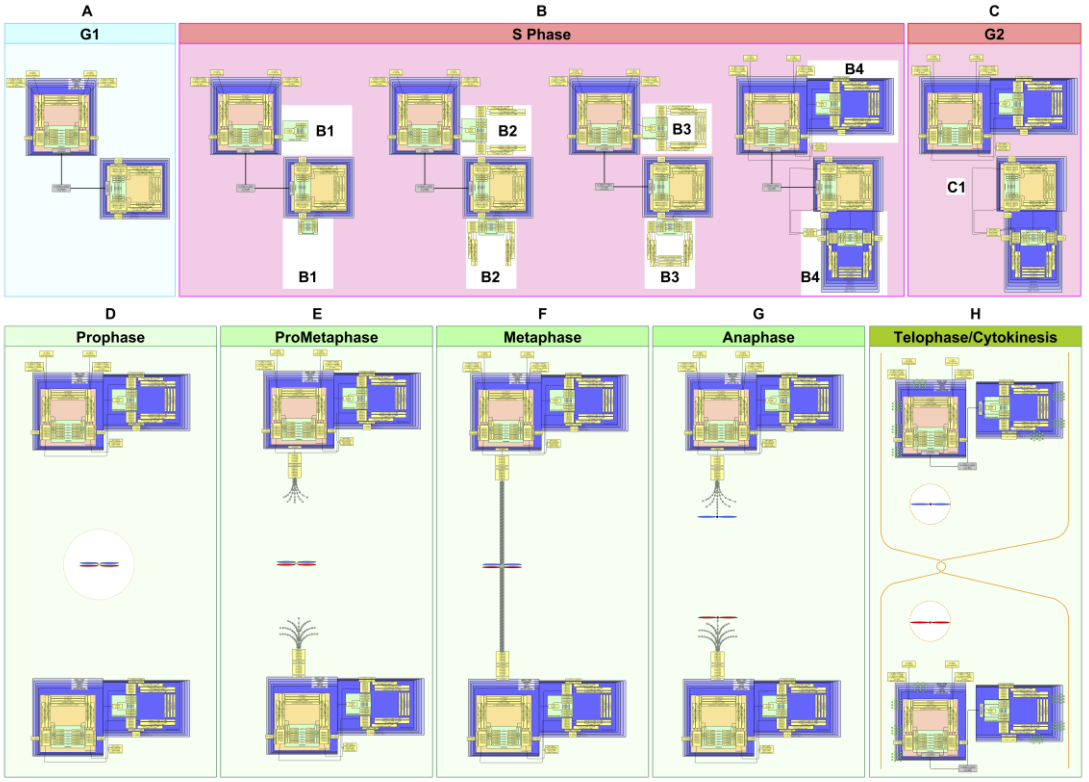


Figure 5.6: Longitudinal section of centrosome duplication at related phases following the progression of cell cycle. A) A matured mother and daughter centrioles paired by linker fibres at G_1-phase. **B)** Interplay of developing centrosomes in S-phase. **B1)** Development of the cartwheel structure. **B2)** Recruitment of the 9-fold symmetrical triple microtubules. **B3)** Recruitment of proteins at distal end of procentriole. **B4)** Redistribution of PCM matures the procentrioles. **C)** Disjunction of a parent centrosome and a daughter centrosome after it is completely matured. **D)** Nuclear envelope break down of a cell and disengaged parental centrosomes during the prophase. **E)** Extension of microtubule assembly. **F)** Attachment of chromosome by microtubules at metaphase. **G)** Microtubule retraction navigates segregated sister chromatids into two daughter cells at anaphase. **H)** Dividing of cytoplasm and chromosomes equally to two daughter cells at telophase/cytokinesis. For full scale images please see attached PDF file in CD.

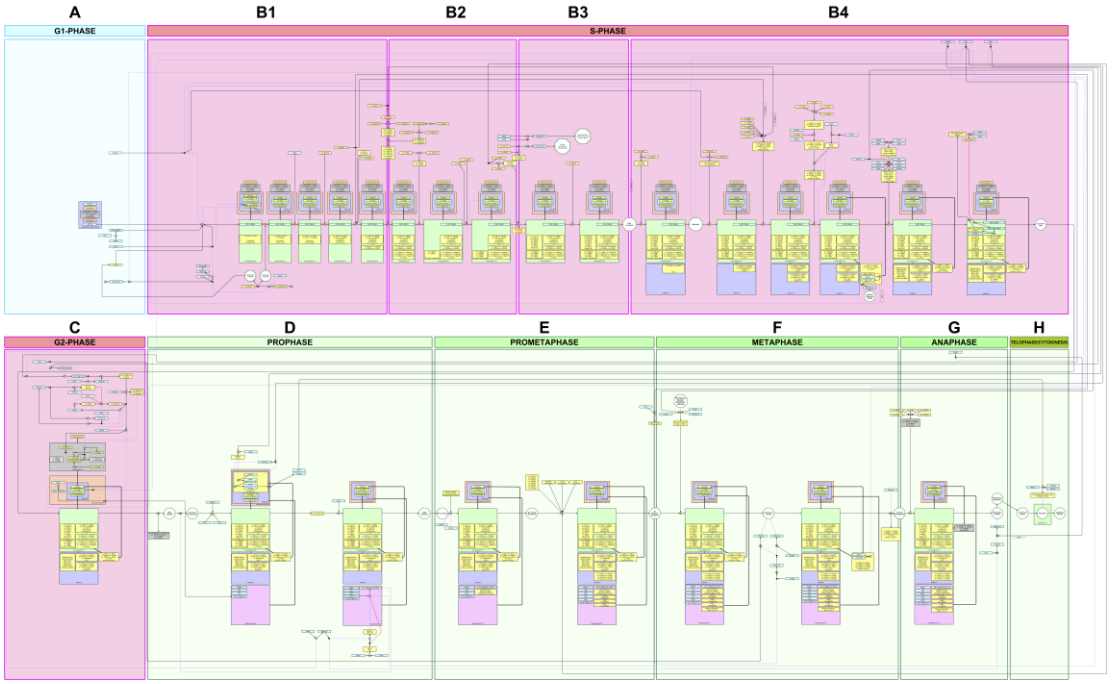


Figure 5.7: Global scale model of a centrosome duplication following a stream of protein-protein interaction network. A) A matured interphase centrosome paired with a daughter centriole by linker fibres. **B1)** Development of cartwheel structures merged from the proximal end of paired parental centrosomes during the S-phase. **B2)** Development of 9-fold symmetrical triple microtubules around cartwheel structures during the S-phase. **B3)** Recruitment of proteins at distal end of procentrioles that against the centriole re-duplication during the S-phase. **B4)** Maturation of procentrioles via the redistribution of PCM from parental centrosomes during the S-phase. **C)** Disjunction of linker fibres between paired parental centrosomes at G_2-phase. **D)** Nuclear envelop break down of cells at prophase and disengaged parental centrosomes heading to opposite positions in a cell. **E)** Extension of microtubules heading to attach sister chromatids during the prometaphase. **F)** Microtubules that attach the kinetochore of aligned sister chromatids during the metaphase. **G)** Microtubule retraction that conduct the chromosomes segregation equally to opposite directions during the anaphase. **H)** Dividing of cells that equally separates the cytoplasm and chromosome to two daughter cells and the reformation of nuclear envelops. For full scale images please see attached PDF file in CD.

Microtubule assembly

Both the mother and daughter centrioles possess a 9-fold symmetry structure of microtubule triplet (Figures 5.5A6 and B1). The long axes of the microtubule determine the length of the cylinder. Each of the microtubule triplets consists of an A-, B-, and C-tubule at the internal, middle, and external positions respectively. In contrast, the length of C-tubule is only two-thirds of the centriole without its distal portion.

The protein A-C linker, which provides the structural connection between A- and C-tubule also maintains the integrity of cartwheel scaffold structure [341]. In addition, the A-C linker connecting the A- and C-tubule belongs to one triple and previous one respectively in the proximal centriole.

PCM

Owing to the essential role of parental centrioles in the progression of the microtubule nucleation, it is important to have PCM, a cloud of proteins, responsible for microtubule nucleation and anchoring activity (Figures 5.5A4 and B5) [52]. During interphase, most of the PCM surrounds the mother centriole, but it is equally distributed around mother and daughter centrioles by the end of G_2-phase after the daughter centriole is completely matured [52]. This ensures that the developing daughter centriole is inactivated during S-phase [52] .

The inner layer functions as a main organiser of the PCM complex (Figures 5.5A4 and B5) [353]. In contrast, the outer layer provides most of the microtubule nucleation functionality by acting as a platform for the docking of γ-TuRC [353].

In order to initiate the assembly of microtubules for segregating the duplicated chromosomes in mitosis, the γ-TuRC functions as the MTOC acting with other centrosome-associated proteins in PCM (Figures 5.6E, F, and G). Only fully developed γ-TuRCs are capable to conduct the disjunction of mitotic chromosomes.

For the microtubule nucleation in M-phase organised by PCM, it is γ-tubulin that plays a pivotal role providing a platform for the initiation of the polymerisation of α- and β-tubulin heterodimers into developing microtubules (Figures 5.6E, F, and G). Collectively, the tubulin gamma 1 (TUBG1) and tubulin gamma 2 (TUBG2) and other associated proteins are known as γ-tubulin complex proteins (TUBGCP2~6) [354-356]. Two copies of TUBG1 together with one copy of tubulin gamma complex associated protein 2 (TUBGCP2) and one copy of tubulin gamma complex associated protein 3 (TUBGCP3) form a tetramer complex called γ-tubulin small complex (γ-TuSC), other γ-tubulin associated proteins tubulin, gamma

complex associated protein 4, 5, and 6 (TUGCPs 4, 5, and 6) are able to be assembled into a high-order complex known as γ-TuRC [356].

Other core elements of the γ-TuRC in association with PCM, such as neural precursor cell expressed, developmentally down-regulated 1 (NEDD1) and mitotic spindle organizing protein 2 (MZT2) are thought to not to have essential role in the γ-TuRC assembly [357-360]. As a PCM-associated protein located at inner and outer layer of the PCM, NEDD1 is known to be independently recruited by γ-TuRC conducting the attachment between γ-TuRC and PCM [358,359,361-363]. Mitotic spindle organizing protein 1 (MZT1) is co-localised with NEDD1 at the inner layer of PCM and is required for the recruitment of γ-TuRC to centrosomes during the M-phase [153,364].

The PCM are a layered structure composed of various proteins from inside to the surface of its structure (Figures 5.5A4 and B5). Of the outer layer proteins in PCM complex, protein pericentriolar material 1 (PCM1) is known to locate at the centriolar satellites that conduct the assembly of centrosomal proteins and microtubule [365]. For the proteins that are equally distributed across inner and outer layers of PCM including a kinase anchor protein 9 (AKAP9) [366] and pericentrin (PCNT) [367], PCNT is known to be the integral component of PCM in the initial establishment of microtubule arrays organisation in both mitosis and meiosis [368]. AKAP9 provides sites for microtubule nucleation in the mammalian centrosome by anchoring γ-TuRC by binding with TUBGCP2 and/or TUBGCP3 at their amino-terminal regions [369].

For the protein complex positioned in inner layer of PCM, CDK5 regulatory subunit associated protein 2 (CDK5RAP2) is important for the binding of γ-TuRC and functions as negative regulator of centriole disengagement maintaining mother and daughter centriole engagement and cohesion (Figures 5.5A4 and B5) [370,371]. Other PCM-associated proteins located in the inner layer, including centrosomal protein 63kDa (CEP63), centrosomal protein 120kDa (CEP120), and centrosomal protein 152kDa (CEP152), are also necessary components in the regulation of γ-TuRC conducting microtubule assembly in M-phase. CEP63 is required for normal spindle assembly and plays a key role in mother-centriole-dependent centriole duplication, through centrosomal recruitment of CEP152 [372]. CEP152 acts as a molecular scaffold facilitating the interaction of polo-like kinase 4 (PLK4) and CENPJ, both of which are involved in centriole formation [373]. Both CEP63 and CEP152 are dependent on one another for the localisation of mitotic centrosome [372]. CEP120 interacts with CENPJ and positively regulates centriole elongation [374].

Distal and subdistal appendages of the mother centriole

Distal appendages promote the membrane anchoring [50]. A complete structure composed of distal and subdistal appendages is an extension network that equally radiates from the distal end of the mother centriole (Figures 5.5A7 and B6) [375]. In distal appendages, protein CEP164 and outer dense fibre of sperm tails 2 (ODF2) are thought to be the conservative elements involved in primary cilia formation (Figures 5.4B and 5.5) [376]. CEP164 is required in the distal appendages of mature centrioles [327]. ODF2 functions as a general scaffold protein in both distal and subdistal appendages [349].

Subdistal appendages conduct the nucleating of microtubule network (Figures 5.5A7 and B6) [51,334]. The subdistal appendage contains protein ninein (NIN), centriolin (CNTRL), tubulin, epsilon 1 (TUBE1), kinesin family member 24 (KIF24), microtubule-associated protein, RP/EB family member 1 (MAPRE1), MAPRE2, MAPRE3, ODF2 and centrosomal protein 170kDa (CEP170) (Figures 5.4B, 5.5A7, and 5.5B6) [376]. NIN is known to be a centrosomal protein required for the position and anchorage of microtubules [377,378]. CNTRL is involved in the centrosome maturation [379]. TUBE1 is involved in the microtubule assembly during centriole duplication [380]. KIF24 is a microtubule-dependent motor protein responsible for mediating recruitment of CCP110 to the mother centriole in cycling cells [381]. MAPRE1, 2, and 3 are known to play important roles in microtubule dynamic regulation, cytokinesis, and mitotic spindle positioning [382]. CEP170 is known to function as the major microtubule-organizing centre [383,384].

The centrosome life cycle

Owing to its complexity and role in the regulation of cell cycle, particularly during S- and M- phase, a categorisation of proteins and their regulation network is formed in this analysis. The following section describes the processes of centriole duplication and maturation during S-, G_2-, and M-phase of the cell cycle by using cartoon-style flow chart and regulatory network in the pathway map (Figures 5.6 and 5.7).

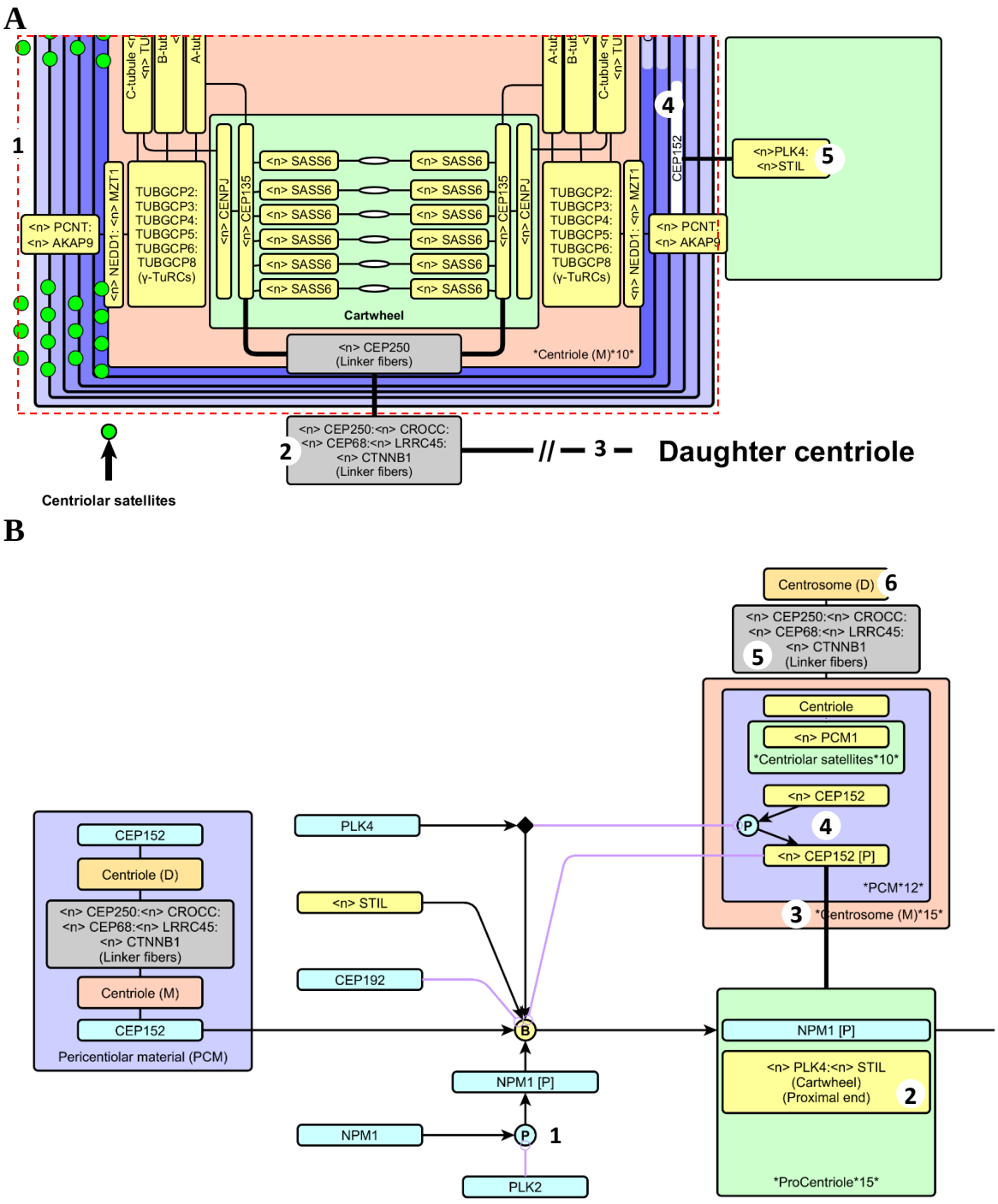


Figure 5.8: Licence activation of NPM1 that initiates the cartwheel structure development. Proteins that located in **A)** the longitudinal section of centrosome structure and **B)** network motif in pathway map are labelled by numbers. **A1** and **B3)** Structure of parental centrosome showed at longitudinal section (red square with dash line) and pathway. **A2** and **B5)** Linker fibres. **A3** and **B6)** Daughter centrioles. **A4** and **B4)** CEP152 located in parental PCM is phosphorylated resulting in the recruitment of PLK4 and SCL/TAL1 interrupting locus (STIL). **A5** and **B2)** Cartwheel structure development is located at proximal end of the parental centrosome. **B1)** Recruitment of phosphorylated nucleophosmin family member 1 (NPM1) which is catalysed by PLK2 in cytoplasm.

Centrosome duplication initiated in G_1-phase

Starting on G_1-phase, the polo-like kinase 2 (PLK2) catalyses the phosphorylation of protein NPM1 shuttling it into the unduplicated centrioles and licenses the development of the cartwheel structure (Figure 5.8B1) [25,385,386]. Phosphorylated NPM1 when initiates centrosome duplication during G_1/S-phase at the mother centriole [387,388], when a number of key proteins are recruited. This includes PLK4 [389], CEP152 [389], CENPJ [371,390-392], and SASS6 [393] initiating the processes of procentriole development (Figure 5.8B). Prior to the development of procentrioles, CEP152 is phosphorylated. This is mediated by PLK4 (Figures 5.8A4 and B4). PLK4 and STIL then localise to the proximal end of the mother centriole attaching on phosphorylated CEP152 where the SASS6 will be recruited. This initiates the cartwheel structure development (Figures 5.8A5 and B2) [394-397]. Additionally, centrosomal protein 192kDa (CEP192) and CEP152 are both involved in the PLK4 recruitment and centriole duplication [398]. For other locations (labelled by numbers), position figure 5.8A1 and B3 represent a longitudinal section and notations represent mother centrosomes in the pathway. Figure 5.8A2 and B5 represent the linker fibres. Figure 5.8A3 and B6 represent the daughter centrioles.

Duplication of the daughter centriole in S-phase

To avoid the over reproduction of procentrioles, cadherin 1 (CDH1) and f-box and WD repeat domain containing 5 (FBXW5) are involved in the ubiquitination of over expressed SASS6 during the development of cartwheel structure (Figure 5.9B2) [340,399]. This double-safety mechanism prevents the re-initiation of procentriole development. STIL then interacts with SASS6 binding to procentrioles stabilising the cartwheel structure (Figure 5.9B1) [400-403]. CEP135 is then recruited to localise nearby the SASS6 at the proximal end (Figure 5.9B3) [404-406]. CENPJ is then recruited to localise beside the CEP135 preparing to nucleate the microtubule (Figure 5.9B4) [371,390-392].

Once the cartwheel assembly is positioned on the proximal end of a mother centrosome, CENPJ and other proteins are then involved in the nucleation and extension of microtubules (Figure 5.10). Following in an orderly fashion, the A-tube of a microtubule is first added to the tail of the radiation spokes of SASS structures binding on γ-TuRC complexes and then continues microtubule assembly of the B- and C-tubes until whole structure is completed for cross and longitudinal section of a cartwheel structure and microtubules of A-, B-, and C-tubes) (Figure 5.5A6 and B1) [407].

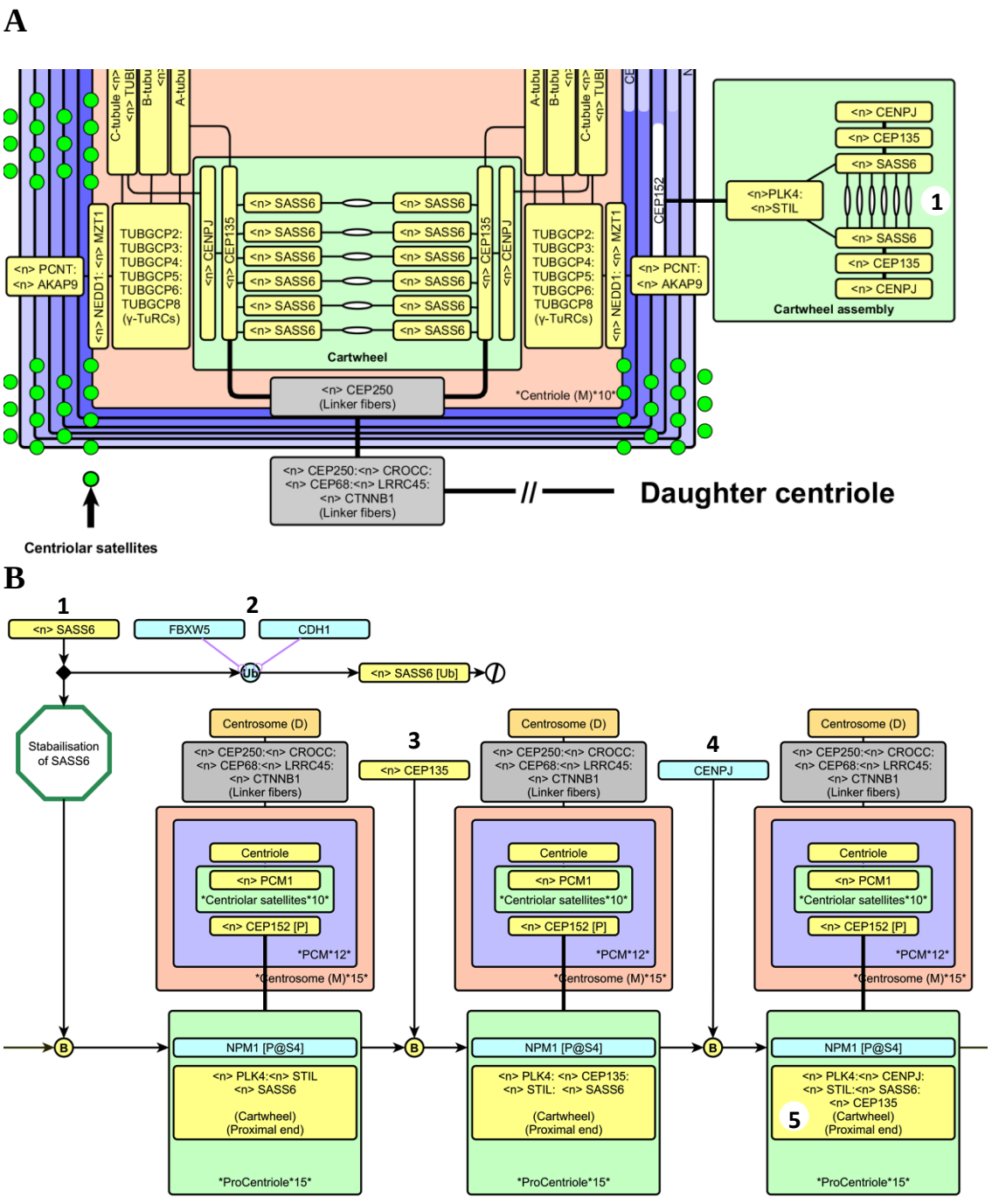


Figure 5.9: Initiation of procentriole development and the emerging of cartwheel structure. A) Longitudinal section of cartwheel structure development. **B)** Pathway of the cartwheel-structured protein assembly. **A1** and **B5)** Location of fully developed cartwheel structure at longitudinal section of procentrioles structure and pathway. **B1)** Expression and stabilisation of SASS6. **B2)** Ubiquitination of over expressed SASS6 regulated by FBXW5 and CDH1. **B3)** Recruitment of CEP135. **B4)** Recruitment of CENPJ.

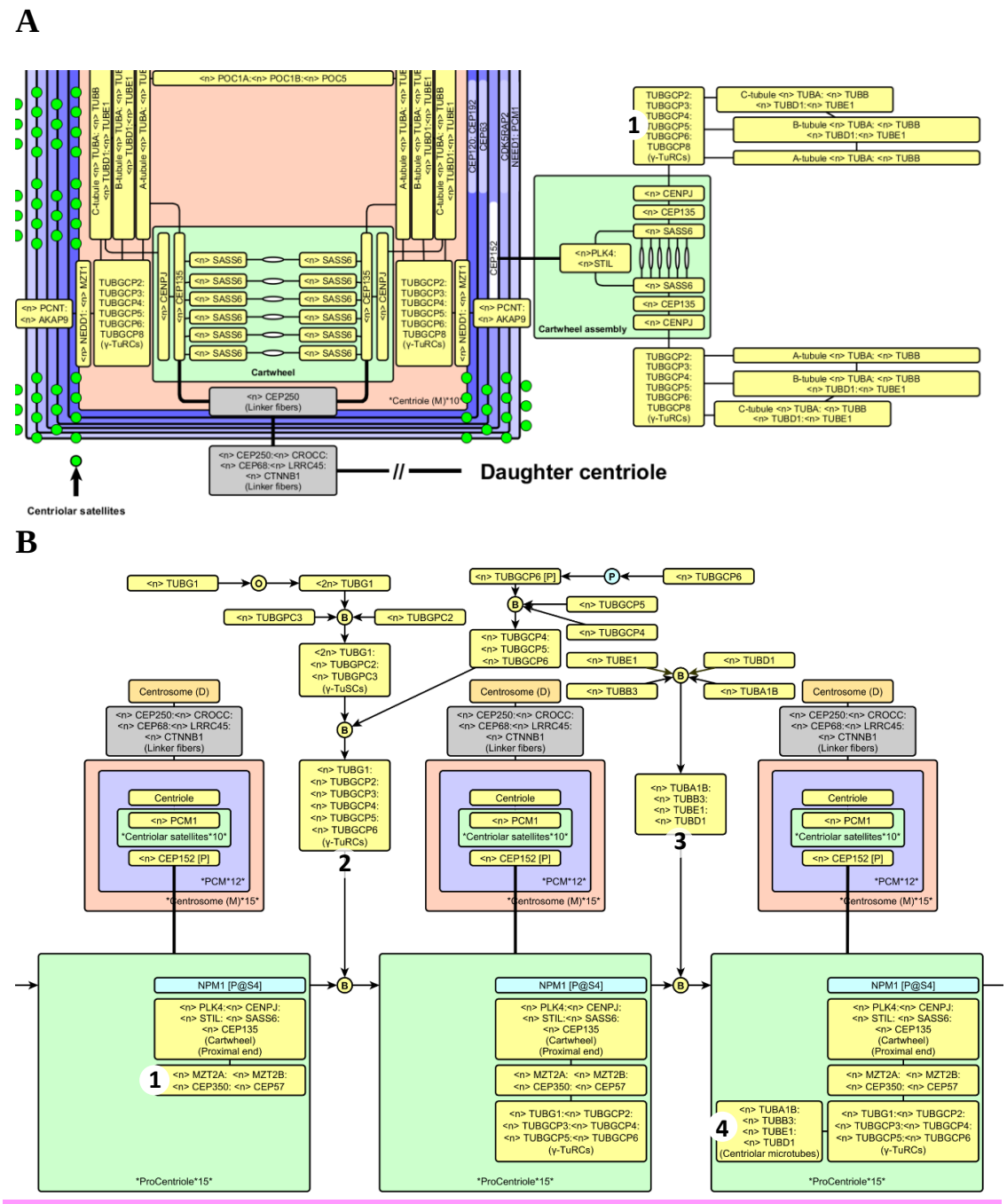


Figure 5.10: Assembling and extension of the microtubule in procentriole development. A) Longitudinal section of microtubule assembly. **B)** Pathway of the microtubule assembly. **A1** and **B4)** Location of fully developed 9-fold symmetrical microtubule triplicate structure at longitudinal section of procentrioles structure and pathway. **B1)** Protein complex, which mediates the recruitment of microtubule. **B2)** Microtubule assembly of γ-TuRC complexes. **B3)** Microtubule assembly of α- and β-tubulins.

Progression of A-, B-, and C-tube recruitment constructs a microtubule-based cylindrical structure, Centrosomal Protein 350kDa (CEP350) is thought to be involved in the developing of procentrioles by stabilising the microtubule structure (Figure 5.10B1) [408]. Protein Centrosomal Protein 57kDa (CEP57), MZT2A, and MZT2B are thought to be involved in the recruiting of γ-TuRC complexes during the interphase (Figure 5.10B1). CEP57 localises around the proximal end of the centriole during the interphase and then conducts microtubule anchoring in M-phase (Figure 5.10B1) [270]. MZT2A and MZT2B are involved in the recruitment of γ-TuRC complexes for PCM of centrioles undergoing maturation (Figure 5.10B1) [360,364].

Microtubule assembly that starts from a template is provided by γ-TuRC complexes (Figure 5.10B) [354,356]. For the development of γ-TuRC complexes, two copies of TUBG1 together with TUBGCP2 and TUBGCP3 are clustered together forming a tetramer called γ-TuSC [409,410]. Other family members of γ-tubulin including TUBGCP4, 5, and 6 are also recruited to bind with γ-TuSC resulting in a higher order of γ-TuRC complexes [356,410-413]. In comparison with the A-tube that is composed of TUBA1B, TUBB3, and γ-TuRC complex, B- and C-tube of microtubules are assembled with the recruitment of additional TUBD1 and TUBE1 (Figures 5.10A1 and B4) [414,415].

The centrin-binding protein POC5 is involved in the regulation of distal procentriole development (Figure 5.11B3) [347]. Other proteins, including CCP110 [381], PCOC1A [346], and POC1B [346], are also recruited to the procentriole structure at distal end (Figures 5.11A1 and B4). Protein USP33 interacted with CCP110 binding to procentriole in late S-phase and finishes the capping of procentriole at distal end (Figure 5.11B2) [416]. In addition, it is interesting that protein Cyclin F (CCNF) is involved in both of the inhibition of over reproduction of DNA and the centrosome by catalysing the ubiquitination of protein CCP110 (Figure 5.11B1) [416]. CETN1 located at the distal end of procentriole during the interphase, plays a role in positioning and segregation of the centrosome in mitosis (Figures 5.11A1 and B5) [352].

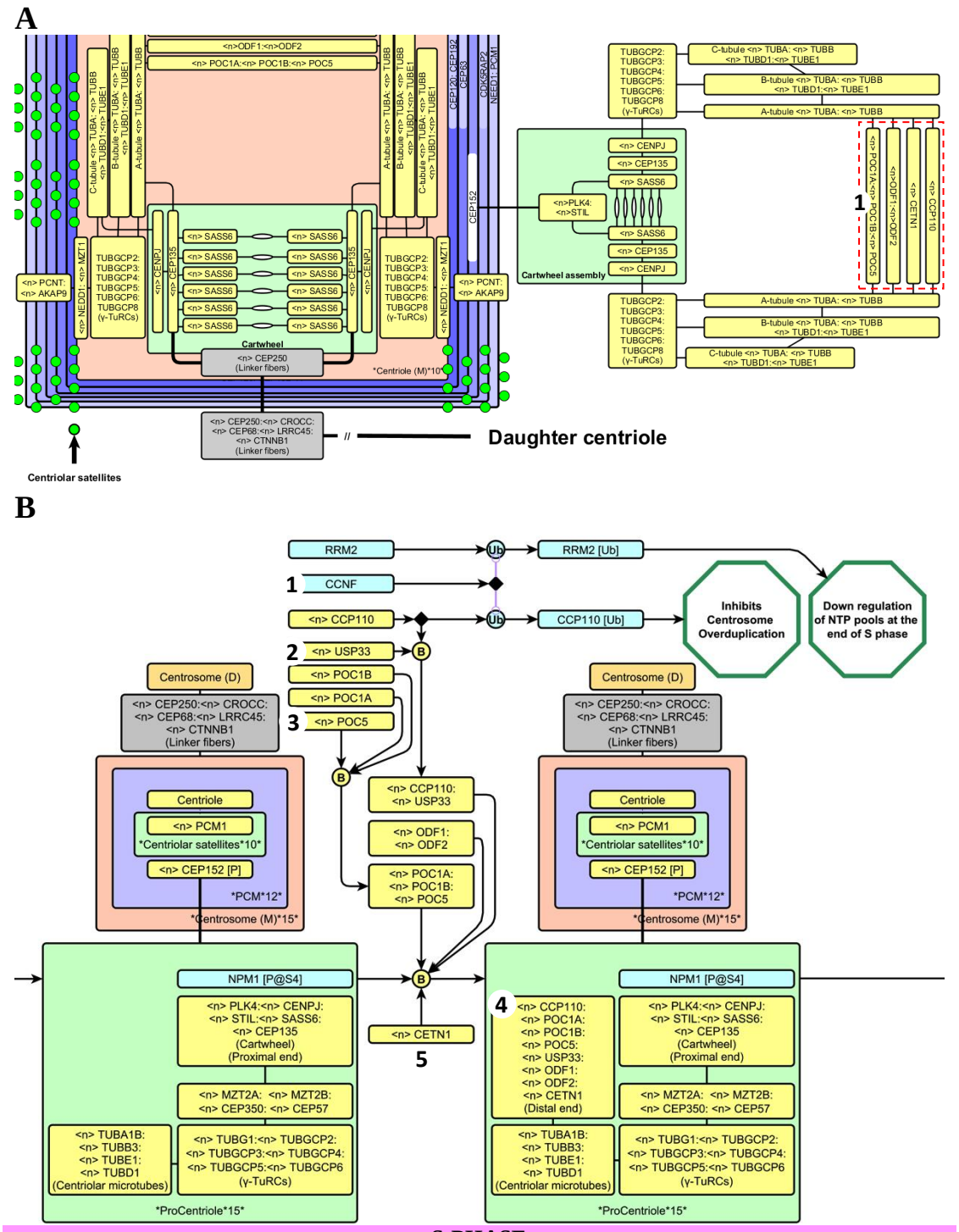


Figure 5.11: Recruitment of the protein complexes at distal end of procentriole. A) Longitudinal section of a procentriole. **B)** Pathway of the recruitment of linker fibres and the proteins at distal end of the procentriole. **A1** and **B4)** Protein complex at distal end of procentriole structure. **B1)** DNA and centrosome reproduction are regulated by CCNF. **B2)** USP33 mediates the recruitment of CCP110 to procentrioles terminating the centriole duplication. **B3)** POC5 regulates the distal procentriole development. **B5)** CETN1 conducts the positioning and segregation of the centrosome.

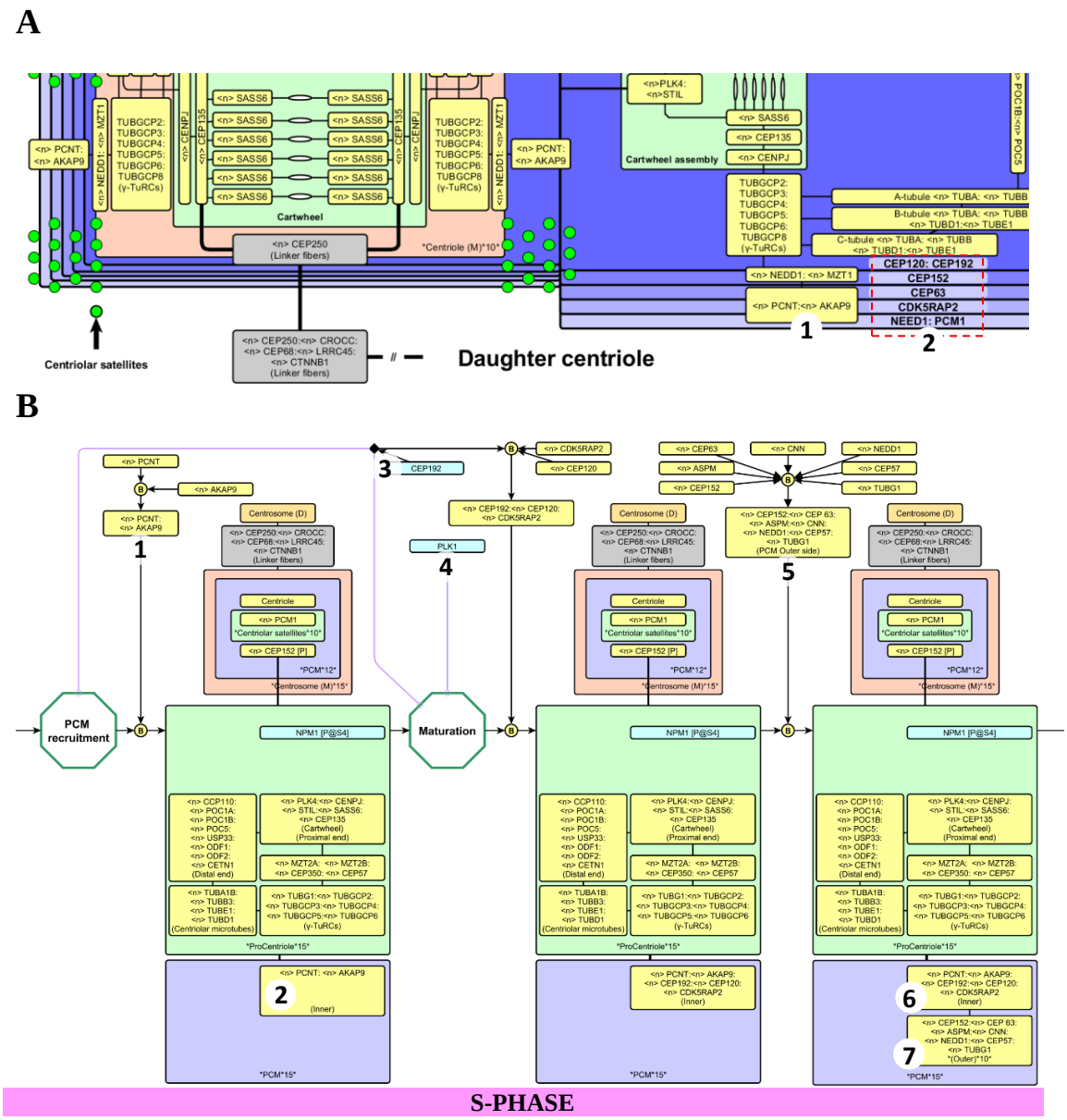


Figure 5.12: Recruitment of PCM and the maturation of daughter centrioles. A) Redistribution of PCM between a parental centrosome and a procentriole. **B)** Pathway of the recruitment of PCM distributed from parental centrosomes. **A1, B1, B2)** PCNT and AKAP9 conducts the recruitment of PCM from mother centrosome from inner to outer layer. **A2, B5, B6, B7)** PCM-associated proteins located at inner and outer layer. **B3)** Located at the inner layer, CEP192 not only recruits CEP120 and CDK5RAP2 to PCM but also catalyses the recruitment, maturation of it. **B4)** PLK1 catalyses the maturation of daughter centriole.

The layered structure of PCM includes inner and outer sphere orbited by centriolar satellites. The recruitment of PCM for procentriole initiates from the nucleating of PCNT and AKAP9 localised at inner layers, which forms the PCM structure (Figures 5.12A1, B1, and B2). In addition, PLK1 induces the centrosome maturation which is essential for procentriole in the next cell cycle (Figure 5.12B4) [417-419]. CEP192 not only is involved in the catalysis of PCM recruitment and maturation but also localises with CEP120 and CDK5RAP2 at the inner layers of the PCM structure (Figure 5.12B3) [416,420]. Clusters of functional protein complexes including NEDD1 [358,359,361], centrosomin (CNN) [52], CEP63 [52], abnormal spindle homolog, microcephaly (ASPM) [52], CEP152 [52], and TUBG1 [52] are located at outer layer of the PCM structure (Figures 5.12A2, B5, B6, and B7).

The disjunction between mother and developing procentriole is known to be a prerequisite to licence the centrosome reproduction and is tightly regulated by PLK1 (Figure 5.13B8) [417]. In order to prevent the premature centrosome splitting during the interphase, cohesion is recruited to stabilise the connection between mother and procentriole (Figures 5.13A1, B3, and B4) [421,422]. In order to against the attacking of separase that induces the disengagement between the mother and developing procentrioles, protein coiled-coil and C2 domain containing 1A (CC2D1A) and sperm associated antigen 5 (SPAG5) catalyse the recruiting of double-strand-break repair protein rad21 homolog (RAD21) to the centrosome inactivating separase (Figures 5.13B1 and B6) [421,423,424].

In addition, in combination with protein phosphatase 2 α (PP2A), cohesin-associated protein shugoshin-like 1 (SGOL1) is recruited to the cohesin complexes being conducted by PP2A (Figures 5.13B2 and B7). The SGOL1 and PP2A complex inhibit the phosphorylation mediated by PLK1 and thereby enables SGOL1 to inhibit the activation of separase (Figure 5.13B5) [425-427].

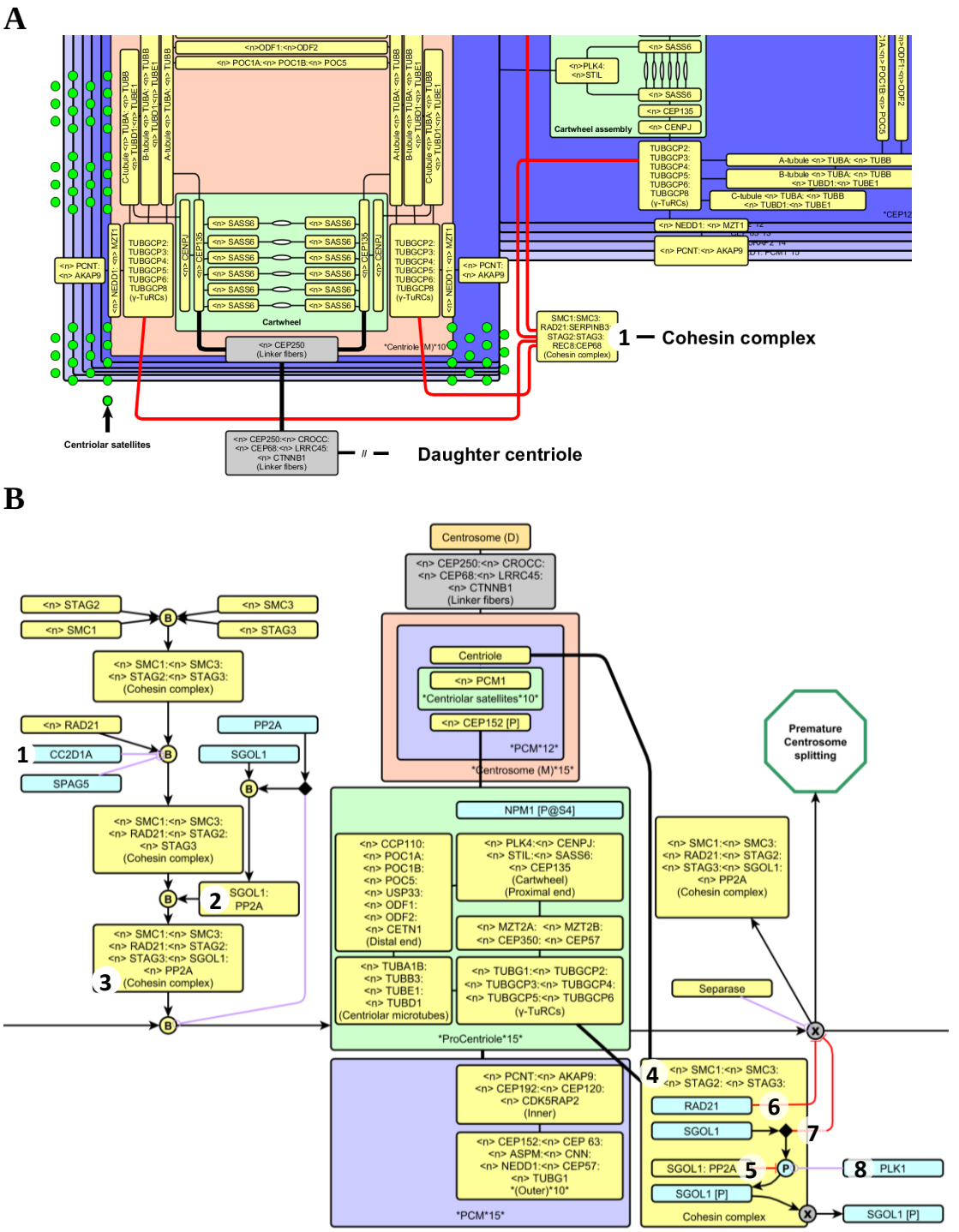
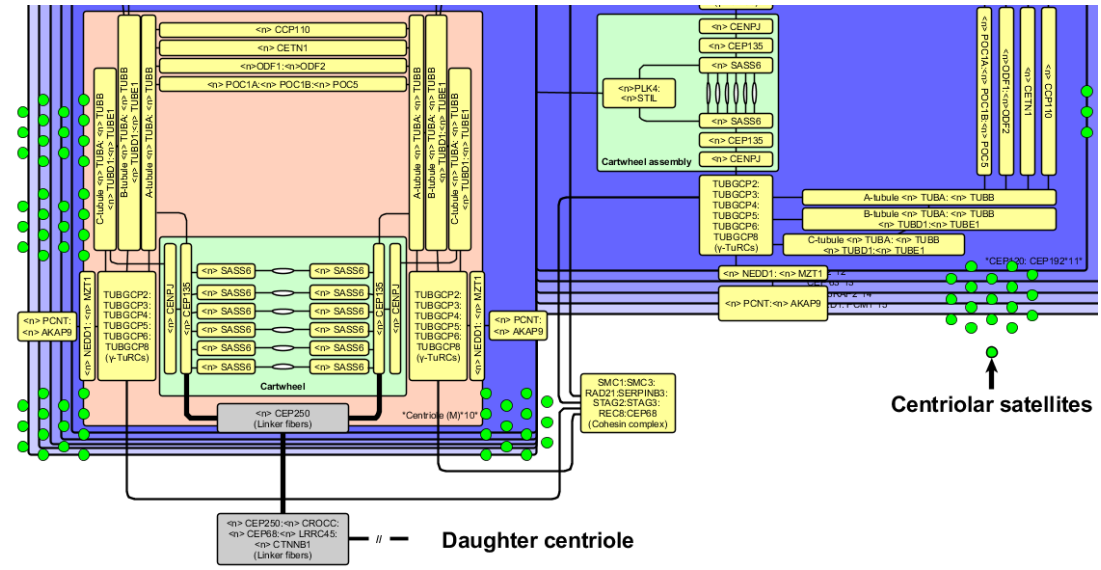


Figure 5.13: Cohesin that engages mother and pre-matured centrosomes is stabilised by SGOL1 and PP2A complexes. A) Cohesin protein complex that tethers paired parental centrosomes at the proximal end (see red line). **B)** Assembly and the recruitment of cohesin-associated protein complex that engages parental and developing centriole. **A1, B3, B4)** Cohesin complex. **B1** and **B6)** Recruitment of RAD21 inhibits the dissociation of cohesin complex. **B2** and **B7)** Recruitment of SGOL1 inhibits the dissociation of cohesin complex. **B5)** A protein complex composed of SGOL1 and PP2A inhibits the phosphorylation of SGOL1 catalysed by PLK1. **B8)** PLK1 mediates the dissociation of SGOL1.

A



B

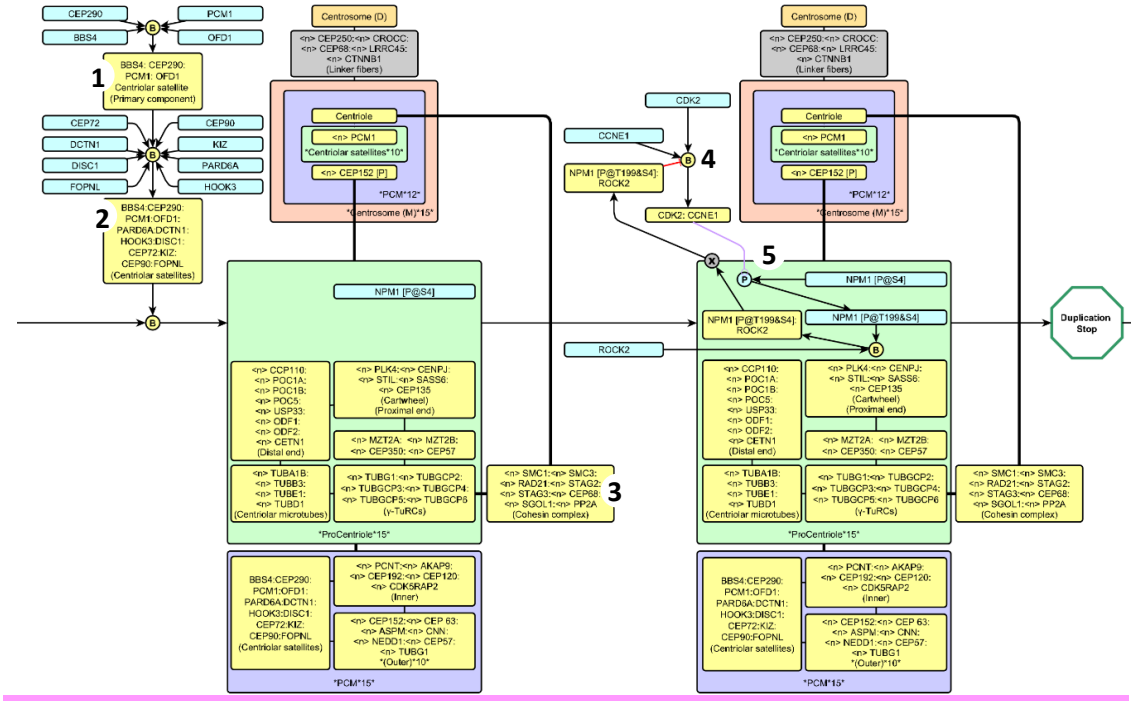


Figure 5.14: Recruitment of centriolar satellites that initiates the nucleating of microtubules. A) Recruitment of centriolar satellites. **B)** Pathway of assembly and recruitment of the cohesin complex-associated proteins. The dissociation of licence proteins that terminates the centrosome duplication by the end of S-phase. **B1)** Recruitment of the primary component of centriolar satellites. **B2** and **B3)** Recruitment of centriolar satellites. **B4)** A protein complex composed of NPM1 and ROCK2 antagonists the binding of CDK2 and CCNE1.**B5)** Dissociation of NPM1 catalysed by the complex composed of CDK2 and CCNE1. For full scale images please see attached PDF file in CD.

The complexes of centriolar satellite-associated proteins are recruited to surround the PCM during the G_2/M border (Figure 5.14B2). The assembly of the centriolar satellite-associated proteins starts from the clustering of primary components, including protein centrosomal protein 290kDa (CEP290), PCM1, bardet-biedl syndrome 4 (BBS4), and oral-facial-digital syndrome 1 (OFD1) (Figure 5.14B1) [428]. CEP290 and BBS4 both cooperate with the PCM1 to be recruited to centriolar satellite [429,430]. OFD1 is involved in the assembly of microtubules for a centriolar satellite [428]. PCM1 is recruited to the centriolar satellite via the cooperation with BBS4 (Figure 5.14B1) [431].

The primary component of centriolar satellite provides a foundation to continue the assembly of other associated proteins including protein dynactin 1 (DCTN1), disrupted in schizophrenia 1 (DISC1), FGFR1OP N-terminal like (FOPNL), hook microtubule-tethering protein 3 (HOOK3), kizuna centrosomal protein (KIZ), OFD1, and partitioning defective 6 α (PARD6A) (Figures 5.14B1 and B2). DCTN1 plays an important role to recruit the centrosome-associated proteins [432] and is also involved in cell division and intracellular transport [433]. In addition, DCTN1 also coordinates with PARD6A and PCM1 to regulate the centrosomal protein recruitment [432]. DISC1 is known to coordinate with BBS4 recruiting PCM1 to the centriolar satellite [431]. HOOK3 coordinates with PCM1 to regulate PCM [434]. FOPNL is localised in the centriolar satellite but there is little information on its biological function [435].

In order to terminate the duplication of centrosome, phosphorylation of licence protein NPM1 is mediated by CDK2 and CCNE1 (Figure 5.14B5) [436,437]. Phosphorylated NPM1 is then combined with ROCK2 and dissociates from the procentriole. The protein complex composed of ROCK2 and phosphorylated NPM1 engages with the combination of CDK2 and CCNE1 that mediates the phosphorylation of NPM1 (Figure 5.14B4) [436,437]. Duplication of centrosome is then terminated after licence protein NPM1 is disengaged from the developing procentriole.

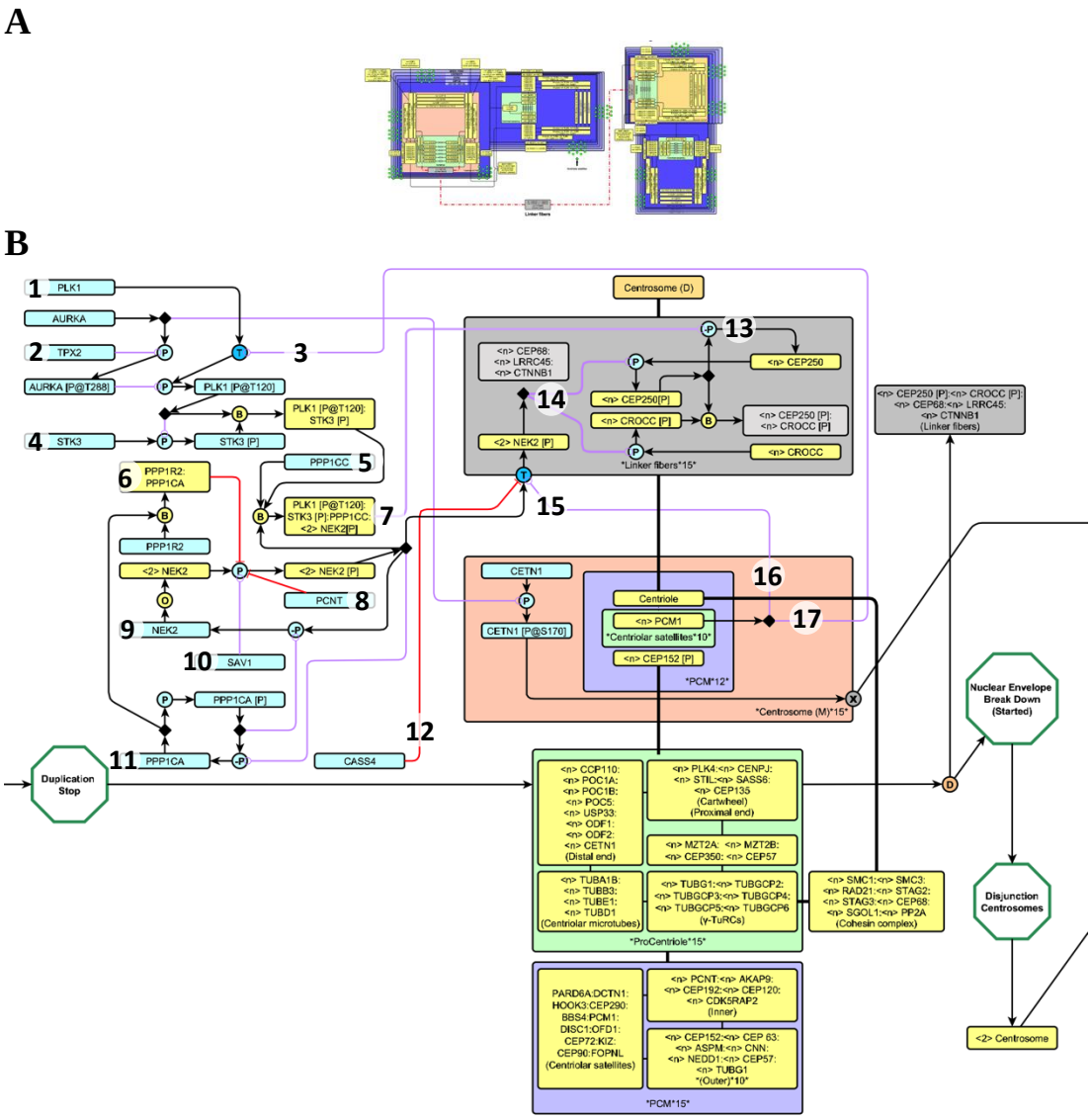


Figure 5.15: Disjunction of mother and daughter centrioles at the G2/M-phase border. A) Dissolution of linker fibres between paired parental centrosomes (see red dashed line). **B)** Regulation of activated kinase that initiates the dissolution of two key linker fibre proteins including centrosomal protein 250kDa (CEP250) and ciliary rootlet coiled-coil rootletin (CROCC) leading to the disjunction between paired parental centrosomes. **B1, B3, B17)** Transition of PLK1 to be phosphorylated by phosphorylated AURKA is catalysed by PCM1. **B2)** AURKA phosphorylation catalysed by TPX2. **B4)** Phosphorylation of STK3 is catalysed by phosphorylated PLK1 and then combined with **B5)** PPP1CC as a protein complex that catalyses the **B7** and **B13)** dephosphorylation of activated CEP250. **B6)** A protein complex composed of PPP1R2 and PPP1CA inhibits the phosphorylation of NEK2 dimers. **B8)** PCNT inhibits the phosphorylation of NEK2 dimers. **B9)** Dimerization of NIMA-related kinase 2 (NEK2). **B10)** Salvador family WW domain containing protein 1 (SAV1) catalyses the phosphorylation of NEK2 dimers. **B15** and **B16)** PCM1 conducts the transferring of phosphorylated NEK2 dimers. **B11)** PPP1CA indirectly conducted the activation and deactivation of NEK2 dimers. **B12)** CASS4 inhibits the transition of activated NEK2 dimers. **B14)** NEK2 dimers catalyse the phosphorylation of CEP250 and CROCC. For full scale images please see attached PDF file in CD.

In order to conduct the segregation of paired parental chromosomes bonded with their mature daughter centrosomes in M-phase, paired parental centrosomes disengage moving to opposite positions in a cell [438]. NEK2 accumulates during the S- and G_2-phase and its activity determines the dissolution of the linker fibres (Figure 5.15B9). NEK2 dimers can be phosphorylated by triggering a down-stream chain reaction to dissolve the linker fibre proteins (Figure 5.15B14) [439,440]. SAV1 is also involved in the catalysis of NEK2 dimer phosphorylation (Figure 5.15B10) [441]. Two key components of the linker fibres, CEP250 and CROCC, are phosphorylated mediated by NEK2 and results in a disjunction of paired parental centrosomes (Figure 5.15B14) [442,443]. The recruitment of NEK2 is conducted by PCM1 from the cytoplasm to localise at linker fibres (Figures 5.15B15 and B16) [444].

As NEK2 plays a deterministic role in linker fibres dissolution, it needs tight regulation in order to prevent a premature centrosome segregation and over reproduction of centrosomes during interphase. An important on-off bi-stable switch at up-stream regulation is the phosphorylation and dephosphorylation of protein phosphatase 1 catalytic subunit α (PPP1CA) that conducts the activation and deactivation of NEK2 dimers (Figure 5.15B11) [445]. Dephosphorylated PPP1CA can also inhibit the phosphorylation of NEK2 dimers by binding with protein PPP1R2 (Figure 5.15B6) [445]. Another inhibitor of NEK2 dimer induced phosphorylation is PCNT (Figure 5.15B8) [446]. CASS4 inhibits the dissolving of linker fibre protein CROCC by preventing the transferring of NEK2 dimers (Figure 5.15B12) [446].

Another branch of the regulation the phosphorylation of aurora kinase A (AURKA) catalysed by TPX2, microtubule-associated (TPX2) (Figure 5.15B2). This is induced by phosphorylated PLK1 trafficking by PCM1 in centriolar satellite (Figures 5.15B1, B3, and B17), which is then combined with phosphorylated serine/threonine kinase 3 (STK3) (Figure 5.15B4), protein phosphatase 1 catalytic subunit (PPP1CC) (Figure 5.15B5), and phosphorylated NEK2 dimer (Figure 5.15B7) [447,448]. This complex antagonises the activation of NEK2 not by directly dephosphorylating but by mediating the dephosphorylation of activated CEP250 to stop the linker fibres dissolution (Figures 5.15B7 and B13) [441].

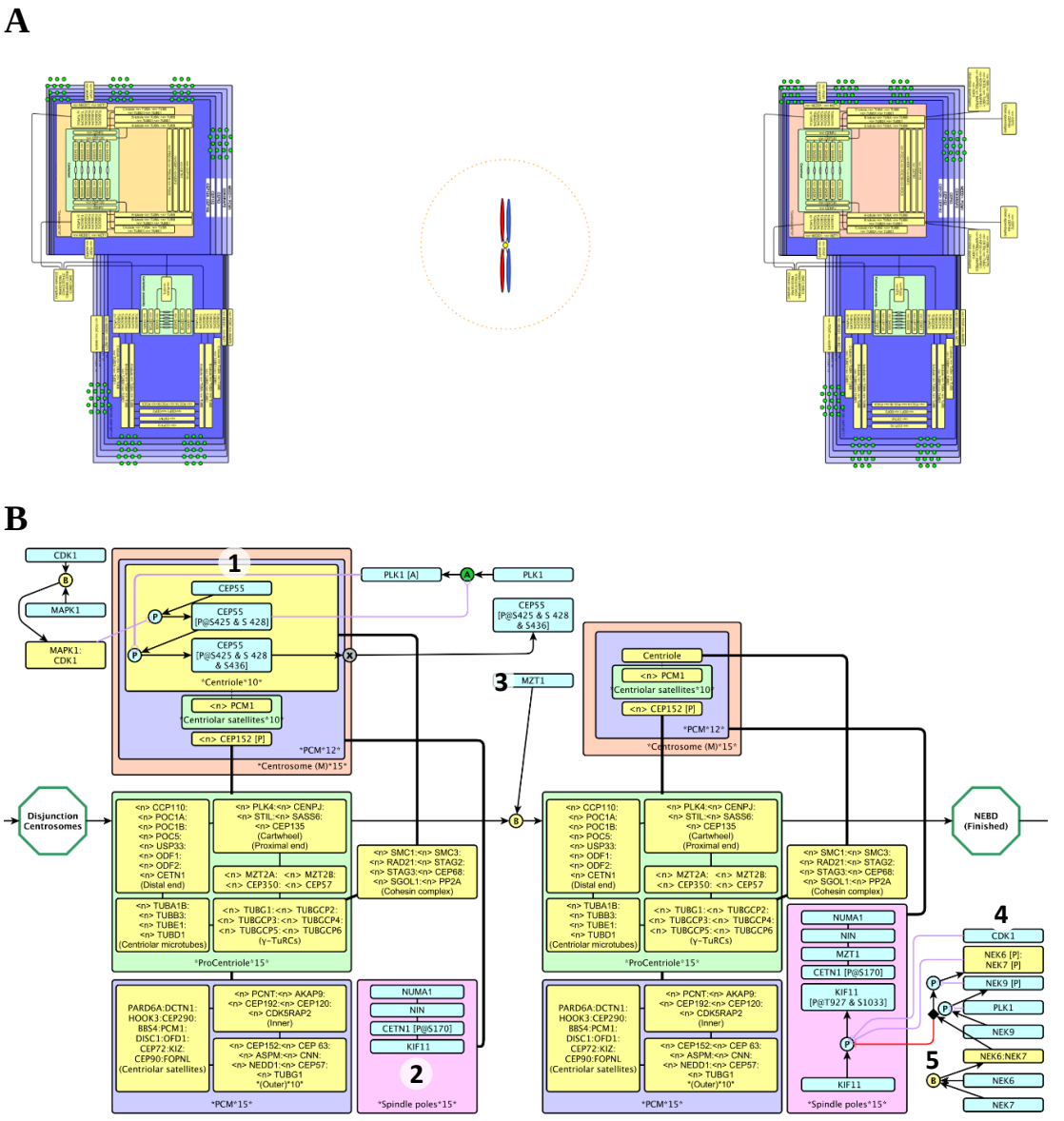


Figure 5.16: Nuclear envelope break-down and recruitment of spindle assembly-regulated proteins. A) Nuclear envelop break down and the separation of disengaged parental centrosomes heading to opposite positions. **B)** Cell cycle-regulated kinases that initiate the dissociation of centrosome protein CEP55 and the regulation of key microtubule assembly-regulated protein KIF11. **B1)** A MAPK1/CDK1 complex initiates the cascade of phosphorylation of CEP55 that dissociates it from the mother centrosome. **B2** and **B3)** Recruitment of protein complex involves in the regulation of microtubule assembly. **B4)** CDK1 and a phosphorylated protein complex composed of NEK6 and NEK7 antagonises the inhibition of a NEK6/NEK7 complex that activates or inactivates KIF11. For full scale images please see attached PDF file in CD.

In addition, AURKA is involved in the phosphorylation of CETN1 in the centrosomes [352,449]. Following the separation of paired parental centrosomes at G_2-phase, the NEBD resulting in two separate centrosomes moving to opposite poles of a cell.

Maturation of the daughter centriole at S/G_2 border

After the disjunction of linker fibres that connect paired parental centrosomes, motor proteins involved in the assembly of microtubules are recruited. To prepare the regulation of microtubule assembly after centrosome separation in M-phase, protein nuclear mitotic apparatus protein 1 (NUMA1) [450], NIN [377,378,451], CETN1 [352], MZT1 [364] and KIF11 are recruited (Figures 5.16B2 and B3). Of these proteins involved in the mitotic microtubule assembly, the kinesin KIF11 plays a major role in centrosome separation and accumulates at spindle poles between G_2- and prometaphase (Figure 5.16B2) [452-454].

Kinase CDK1 and PLK1 determine the activation of KIF11 by directly phosphorylating KIF11 and trigging a cascade of phosphorylation stream composed of NEK9, NEK7, and NEK6, respectively (Figure 5.16B4). To prevent the over-accumulation of activated KIF11 that induce the over extension of microtubules, the complex composed of NEK6 and NEK7 antagonises the activation of the phosphorylated complex (Figure 5.16B5) [46]. The NEBD is finished at the end of prometaphase while disjunctive parental centrosomes complete the recruitment of the regulatory proteins involved in the microtubule assembly.

In combination with MAPK1, CDK1 is then able to catalyse the phosphorylation of CEP55 located in the mother centrosome, which then activates PLK1 catalysing further phosphorylation of CEP55. The phosphorylated CEP55 is then dissociated from the mother centrosome, shifting to midbody conducting the physical dividing of cells during the cytokinesis (Figure 5.16B1) [43-45].

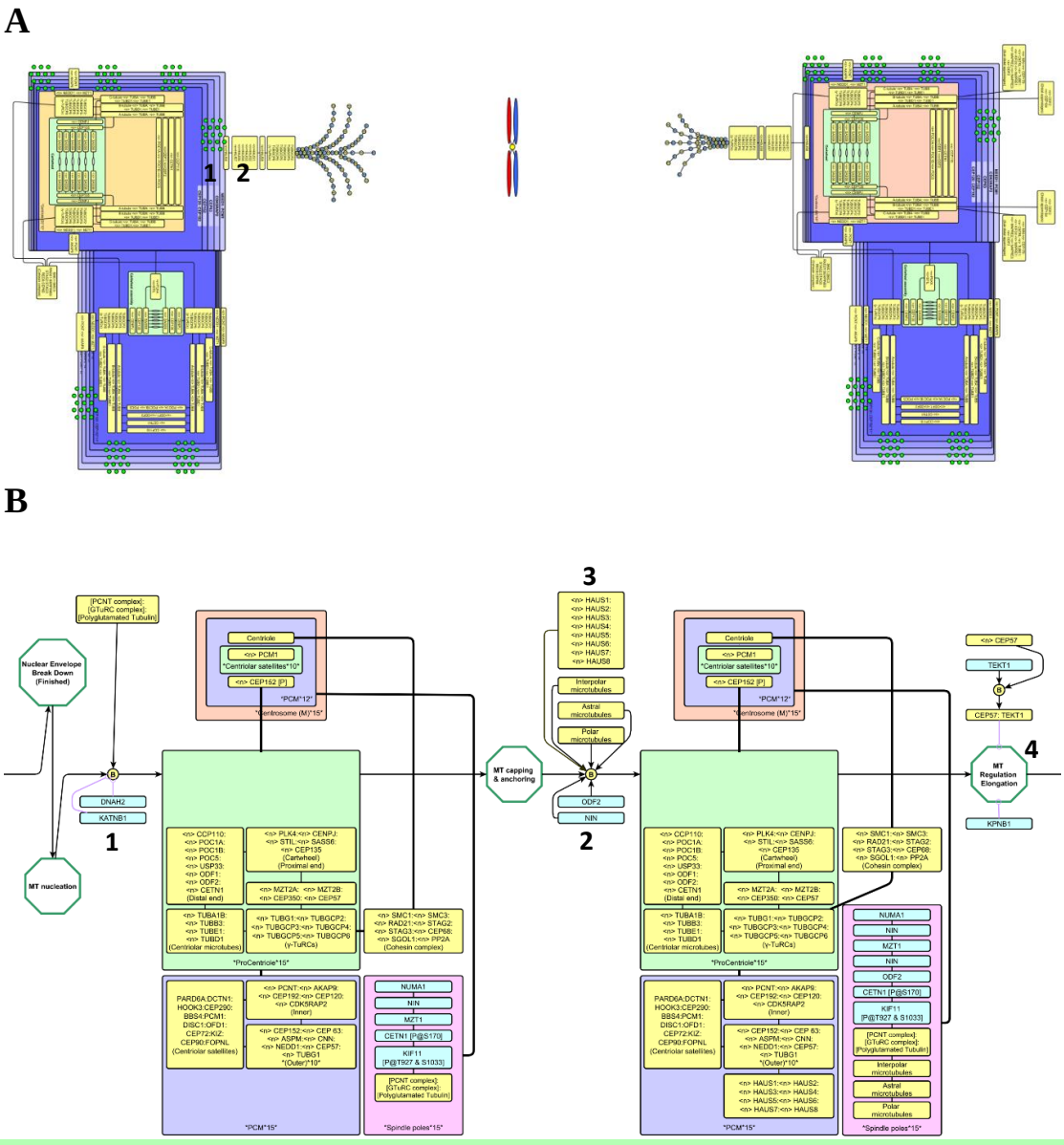


Figure 5.17: Nucleating, anchoring, extension of mitotic microtubule assembly. A) Assembly and extending of microtubule to attach sister chromatids. **B)** Pathway of the recruitment of microtubule-associated protein complexes leading to the elongation of microtubules. **A1** and **B2)** Proteins ODF2 and NIN located at PCM and regulate its recruitment **A2** and **B3)** HAUS family for microtubule capping and anchoring. **B1)** Proteins DNAH2 and KATNB1 catalyse the recruitment of a PCNT complex for microtubule nucleation. **B4)** KPNB1 and CEP57/TEK1 complex catalyse the microtubule elongation heading to the kinetochore of duplicated sister chromatids. For full scale images please see attached PDF file in CD.

The recruitment of PCNT complex is conducted by dynein axonemal heavy chain 2 (DNAH2) and katanin p80 subunit B1 (KATNB1) initiating the microtubule nucleating (Figure 5.17B1). At the stage of microtubule capping and anchoring, interpolar, astral, and polar microtubules are recruited and anchored to the PCM. Protein ODF2 and NIN are recruited to mitotic spindle which is anchored to the PCM of the centrosomes (Figures 5.17A1 and 5.17B2). ODF2 may be involved in the maturation event of daughter centrioles and is required for the development of appendages that render a platform anchoring the microtubules [349,455-457]. NIN is responsible for the anchorage of microtubules to centrosomes [368,377,378]. The family of proteins HAUS1 to 8 are preferentially associated with the γ-TuRC complexes in mitosis and localise γ-TuRC complexes to mitotic spindles (Figures 5.17A2 and B3) [153,358,359]. Protein CEP57 directly recruits and cooperates together with tektin 1 (TEKT1) to trigger the initiation of microtubule elongation localising at central spindle as a spindle organiser in cytokinesis [458]. Karyopherin β1 (KPNB1) is involved in the microtubule assembly in mitosis (Figure 5.17B4) [459,460].

Before attaching of extended microtubules to the aligned chromosomes in the metaphase, a number of proteins involved in the nuclear distribution pathway are recruited to centrosome including protein DCTN1 [334,433,461], NDEL1 [462,463], DNAH2 [464], and platelet-activating factor acetylhydrolase 1B regulatory subunit 1 (PAFAH1B1) (Figure 5.18B1) [464]. These cytoplasmic dynein-regulated proteins are involved in the organization of the cellular microtubule array and microtubule, anchoring at the centrosomes [409].

At metaphase, aligned chromosomes attached by microtubules are transferred to two daughter cells. The binding between the kinetochore of chromosomes and the microtubules are conducted by cascade of the activated NUMA1 and TPX2 triggering by activated KPNB1. TPX2 (i.e microtubule-associated protein) is involved in the regulation of spindle assembly [422,465,466]. KPNB1 plays a role in the microtubule assembly [459,460]. Little is known of the mechanism that actives the KPNB1. NUMA1 is involved in the regulation of microtubules during the mitosis (Figures 5.18A1 and B2) [450].

At telophase, sister chromatids attached by microtubules are disengaged after the dissolution of cohesion being induced by separase. The cohesin that tethers mother and matured sister centrosomes are also dissociated by separase and the phosphorylation of RAD21, which is mediated by PLK1 (Figure 5.18B3) [422].

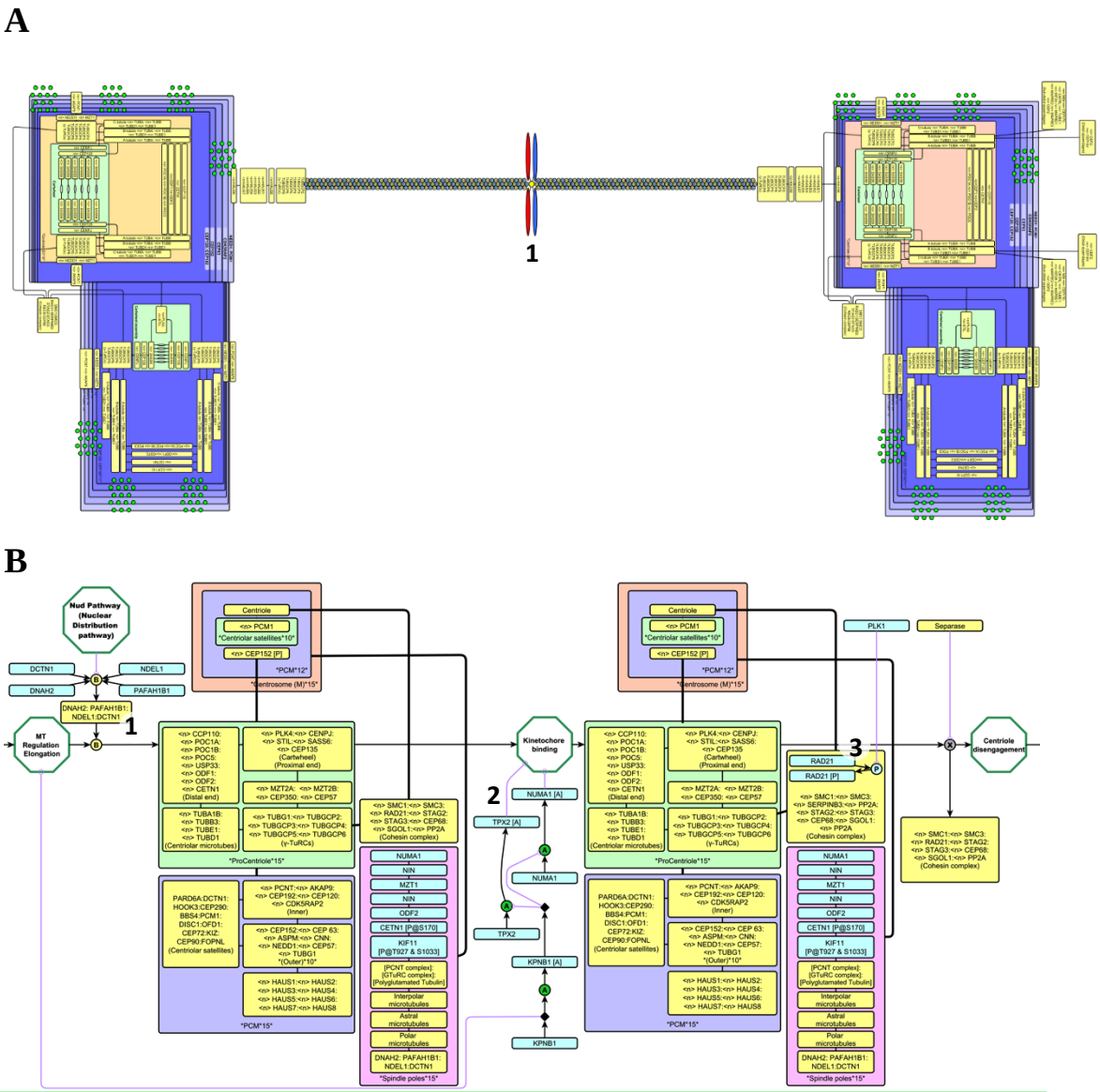


Figure 5.18: Segregation of sister chromatids conducted by mitotic centrosomes. A) Attachment of aligned sister chromatids during the metaphase. **B)** Pathway of nuclear distribution, kinetochore attachment, and the dissociation of cohesin complex between paired parental centrosomes. **A1** and **B2)** Activation of KPNB1 triggers the activation of TPX2 and NUMA1 that drives the kinetochore binding. **B1)** Recruitment of DCTN1, NDEL1, DNAH2, and PAFH1B1 conducts the microtubule elongation toward the attachment of sister chromatids. **B3)** Separase conducts the dissociation of cohensin complex after the phosphorylation of RAD21 catalysed by PLK1. For full scale images please see attached PDF file in CD.

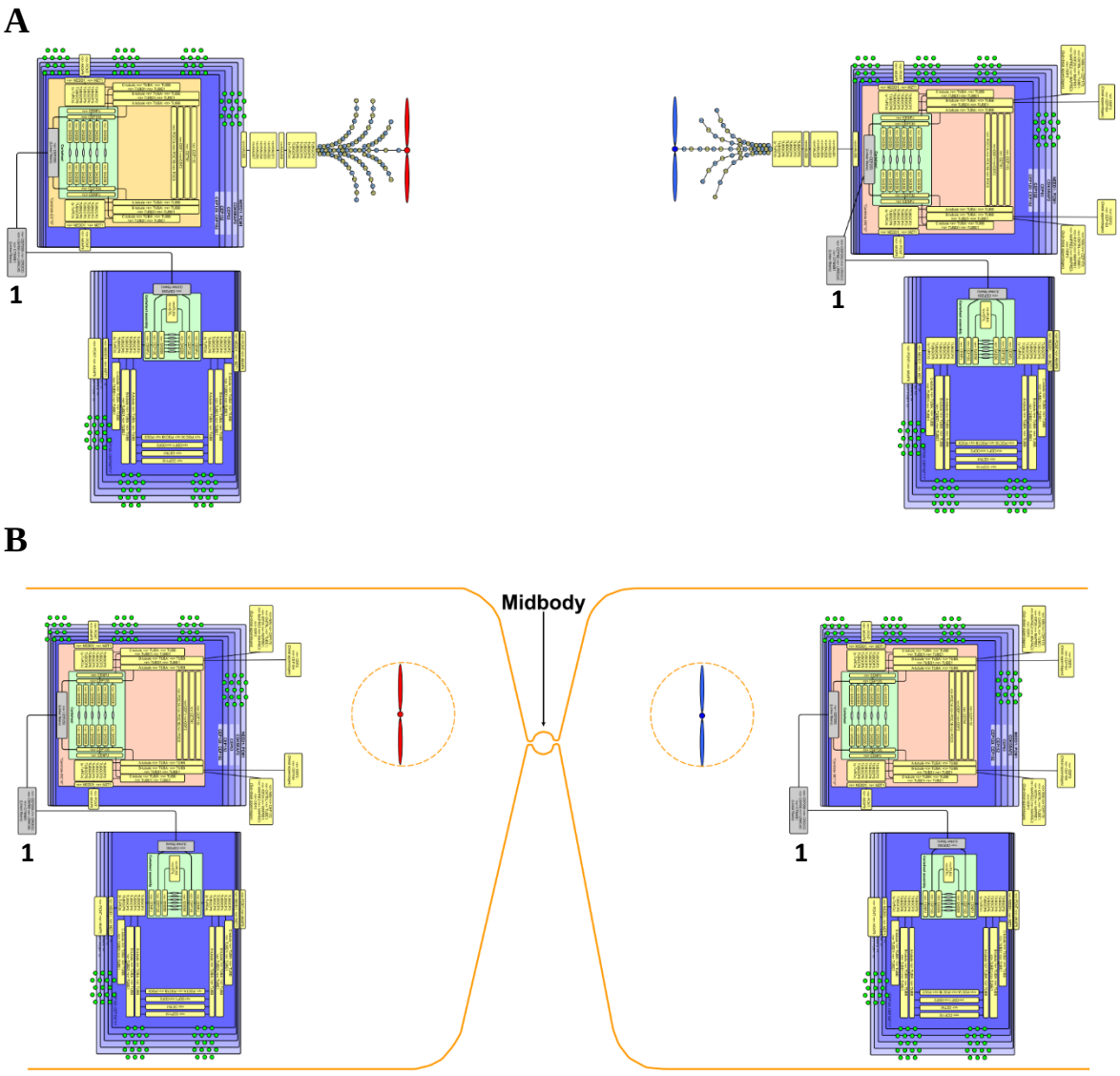


Figure 5.19: Cytokinesis and segregated sister chromatids that are equally transferred into two daughter cells. A) Segregated sister chromatids are equally shifted to two daughter cells being conducted by microtubule retraction at the anaphase. **B)** Reconstructing on the nuclear envelops in two daughter cells at the end of cytokinesis after abscission in midbody. **A1** and **B1)** Linker fibres. For the cytokinesis at system level, please see next page. For full scale images please see attached PDF file in CD.

In Figure 5.19, the linker fibre-associated proteins are then recruited with cohesin to replace the mother and daughter centrosomes in preparation for further centrosome duplication in divided cells (Figures 5.19A1, B1, and C1). Protein fizzy/cell division cycle 20 related 1 (FZR1) is known to function as a key regulator of anaphase promoting complex (ANAPC) [467]. To enter the late stage of M-phase, phosphorylated NEK9 activates the FZR1, leading to the activation of ANAPC that conducts the microtubule retraction (Figure 5.19C2). Disengaged sister chromatids are then shifted to two daughter cells conducting by

microtubule retraction entering the cytokinesis where it physically divides a parental cell into two daughter cells.

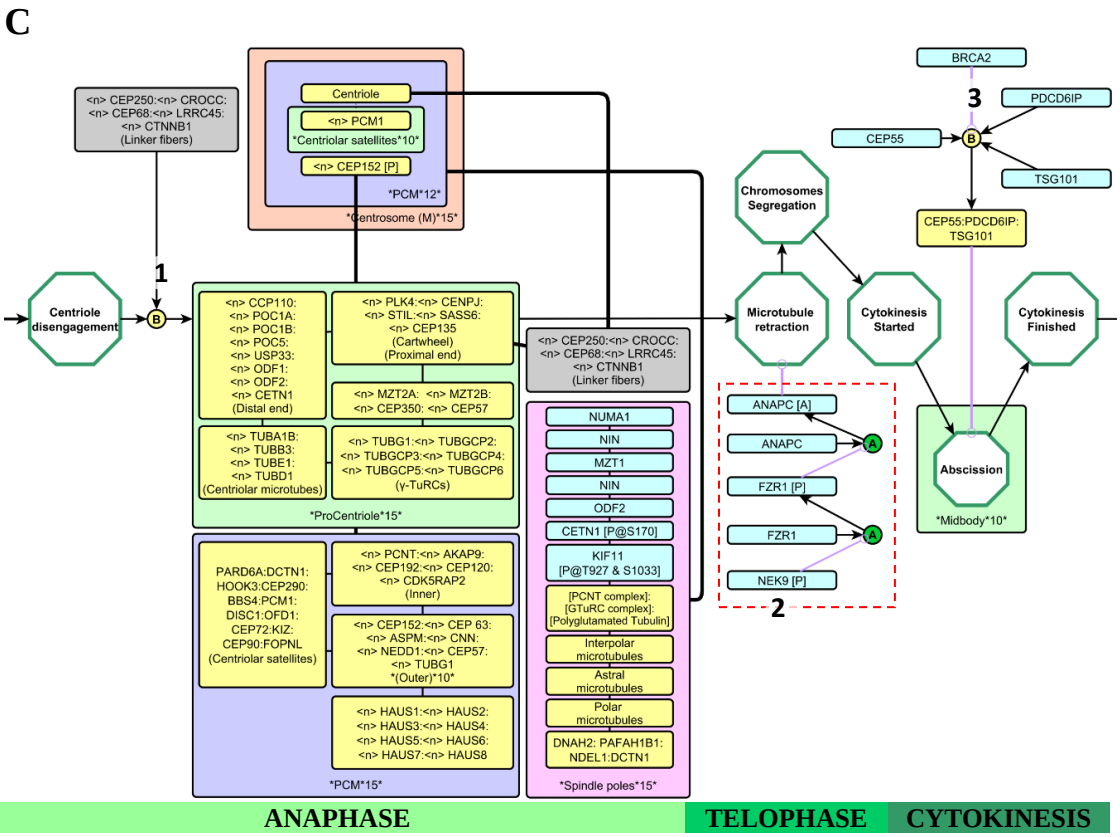


Figure 5.19: Cytokinesis and segregated sister chromatids that are equally transferred into two daughter cells. C) Pathway of the recruitment of linker fibres and the regulation of abscission in midbody at the end of cytokinesis (see next page). **C1)** Recruitment of linker fibres after the dissociation of the cohesion complex. **C2)** Cascade of activation includes FZR1 and ANAPC that triggers the microtubule retraction bring attached chromosomes to the opposite pole. **C3)** BRCA2 catalyses the binding of CEP55, programmed cell death 6 interacting protein (PDCD6IP), and TSG1 conducting the abscission to physically separate cells by the end of the cytokinesis.

To finish the last stage of cytokinesis, abscission conducts the separating of midbody that connect two dividing daughter cells [416]. The midbody provides a platform to regulate the cytoskeleton rearrangement, cell membrane remodelling, and other functional protein complexes recruited for abscission [416]. Protein BRCA2 mediates the recruitment of CEP55, PDCD6IP, and tumour susceptibility 101 (TSG101) to localise at midbody serving to mediate the abscission and finish a complete cell dividing (Figure 5.19 C3) [416].

Expression periodicity of centrosome-associated genes in the pathway map

To further investigate the regulatory mechanism that conducts the centrosome duplication and other associated events at transcriptional level, genes which are overlapped between array dataset and centrosome pathway map were highlighted by colours showing their expression periodicities at different stages of the cell cycle i.e. genes which are consistently expressed and up-regulated during G_1, S/G_2/M-phase (Figure 5.20A).

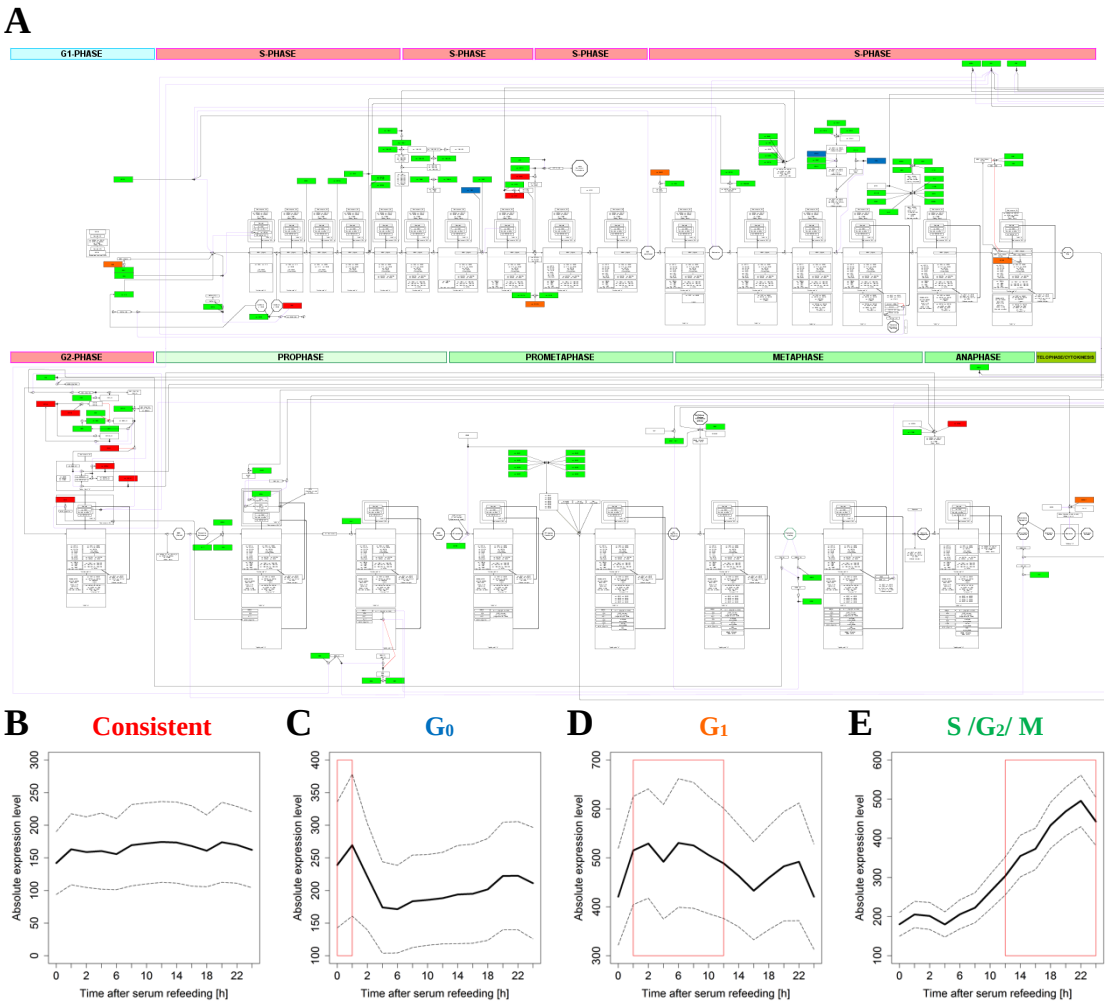


Figure 5.20: Expression periodicity of genes in association with centrosome biology at system level. A) Expression profile of genes that are periodically up-regulated or consistently expressed were labelled by different colours according to the array dataset in this thesis. **B)** Expression level of genes that consistently expressed throughout the cell cycle. **C)** Expression level of genes up-regulated during G_0-phase. **D)** Expression level of genes up-regulated during G_0-phase (red box). **E)** Expression level of genes up-regulated during S-, G_2-, and M -phase. Black dashed line represents variation in gene expression signal. Details of centrosome regulation at transcriptional and proteomic level were summarized in appendix 5.1. Full-scale PDF files are saved in the CD attached. Genes in association with the appendages are only shown in centrosome pathway map at structural level.

Centriole		
Procentrioles	G_1 :	ROCK2
	S/ G_2 /M :	NPM1
Proximal end (Cartwheel structure)		
	S/ G_2 /M :	CENPJ, CEP135, SASS6
Distal end		
	Consistent expression :	CETN1, ODF1, USP33
	G_1 :	POC1B
	S/ G_2 /M :	CCNF, CCP110, POC1A, POC5
9 symmetrical microtubules		
	Consistent expression :	TUBGCP2, TUBGCP6
	G_0 :	TUBE1
	S/ G_2 /M :	TUBA1B, TUBB3, TUBD1, TUBG1, TUBGCP3, TUBGCP4, TUBGCP5
Appendages	G_0 :	MAPRE2
	G_1 :	MAPRE3, CEP170
	S/ G_2 /M :	ODF2, CEP164, CNTRL, MAPRE1, NIN, KIF24
PCM		
Inner layers		
	G_1 :	AKAP9
	S/ G_2 /M :	CDK5RAP2, CEP120, PCNT
Outer layers		
	Consistent expression :	CEP63
	S/ G_2 /M :	ASPM, CEP152, CEP57, CNN, NEDD1
Basement of daughter centriole		
	G_1 :	PLK2
	S/ G_2 /M :	CEP192, PLK4, STIL
Basement of mitotic spindle assembly		
	Consistent expression :	PAFAH1B1
	G_1 :	NUMA1
	S/ G_2 /M :	HAUS5, HAUS1, HAUS2, HAUS3, HAUS4, HAUS6, HAUS7, HAUS8, KIF11
Centriolar satellite		
	Consistent expression :	DCTN1
	G_1 :	DISC1, BBS4, HOOK3, OFD1, KIZ, CEP90
	S/ G_2 /M :	PARD6A, CEP290, CEP72, FOPNL, PCM1
Linker fibre		
	Consistent expression :	CROCC, CEP250, CEP68, LRRC45
	G_1 :	CTNNB1
Cohesin complex		
	Consistent expression :	STAG3
	G_0 :	PP2A, CC2D1A
	S/ G_2 /M :	RAD21, SGOL1, SMC1, SMC3, SPAG5, STAG2
Cytoplasm		
	Consistent expression :	CASS4, CDH1, PPP1CA, PPP1R2, DNAH2, TEKT1
	G_1 :	NEK6, NEK9, NDEL1, SAV1, MAPK1, CCNE1
	S/ G_2 /M :	ANAPC, CDK1, FZR1, KATNB1, KPNB1, PLK1, TPX2, FBXW5, AURKA, NEK2, NEK7, PPP1CC, STK3, CDK2
Midbody		
	Consistent expression :	TSG101
	G_1 :	PDCD6IP
	S/ G_2 /M :	BRCA2

Table 5.2: Expression periodicities of proteins in the pathway map. The expression periodicities of genes relevant to the centrosome pathway map are categorised according to the centrosome structure. Genes highlighted in red (gene that inhibits the centrosome reduplication), blue, orange, green, and black (consistent expression but has not yet been proved to be involved in the inhibition of centrosome reduplication) show their position and expression periodicities in centrosome biology. Genes in association with the appendages are only shown in centrosome pathway map at structure level (Figures 5.5A7 and B6).

Centrosome regulatory network translated in 3D environment

BioLayout *Express*^{3D} is a software using network based analysis utilising 3D graphic application programming interface (API) OpenGL [112]. BioLayout *Express*^{3D} provides a powerful tool to visualise various type of 'omic' data as networks [112]. mEPN diagram drawn in yEd can now be integrated into many tools.

Using BioLayout *Express*^{3D}, a 2D pathway, which is composed of nodes and edges can be viewed in 3D environment. In 3D environment, node walks indicate the parents or children nodes allowing the discovery of network connectivity between components.

The centrosome pathway map in structure and system level was visualised in BioLayout *Express*^{3D}. At system level, a snapshot of centrosome pathway map in 3D shows the nodes, edges, and components representing different functional protein complexes (Figure 5.21) and the centrosome during interphase and M-phase are also visualised in 3D environment at the structural level (Figure 5.22). The centrosome pathway maps at structure and system level are showed in parallel (Figure 5.23).

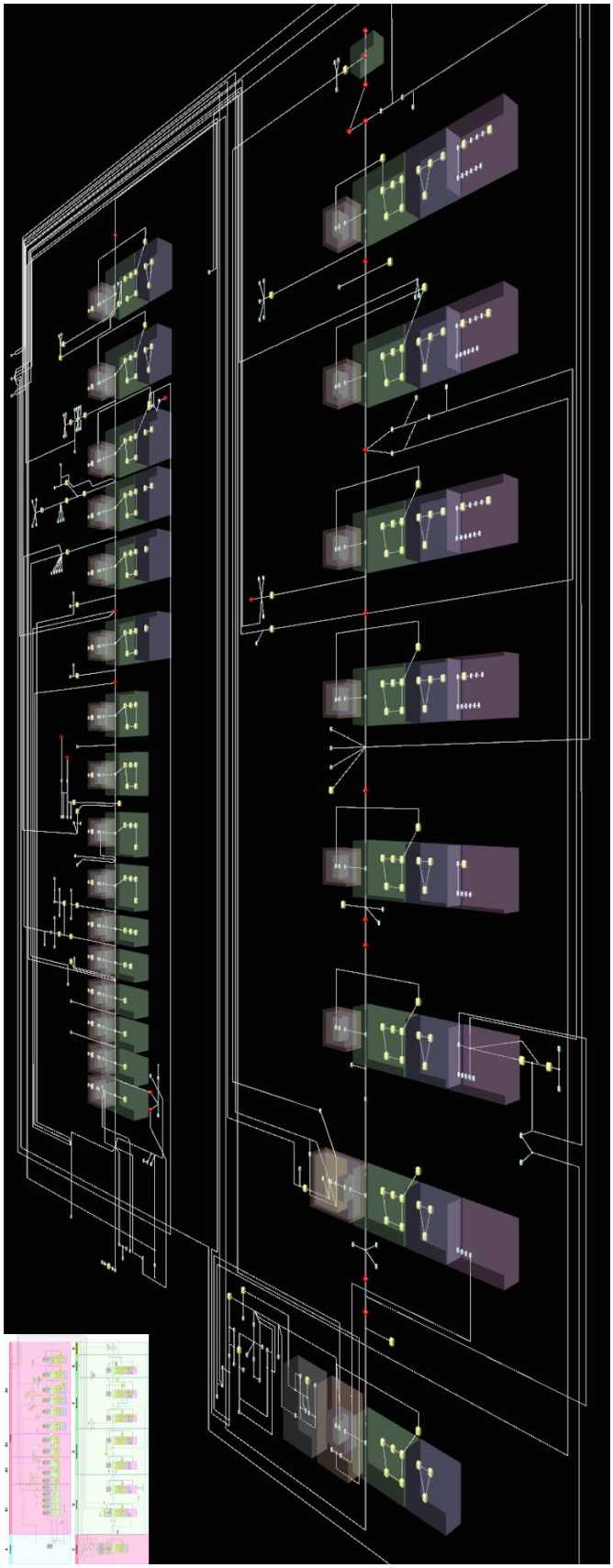


Figure 5.21: Pathway map of centrosome regulatory network translated in 3D environment. Upper left side of the figure shows the pathway in 2D depicted by mEPN in yEd software. Pathway map of centrosome regulatory network depicted in 2D by mEPN scheme was saved as '.graphmal' format and visualised in 3D by BioLayout *Express*^{3D}. For full-scale image, please use BioLayout *Express*^{3D} to explore the pathway map file attached in CD.

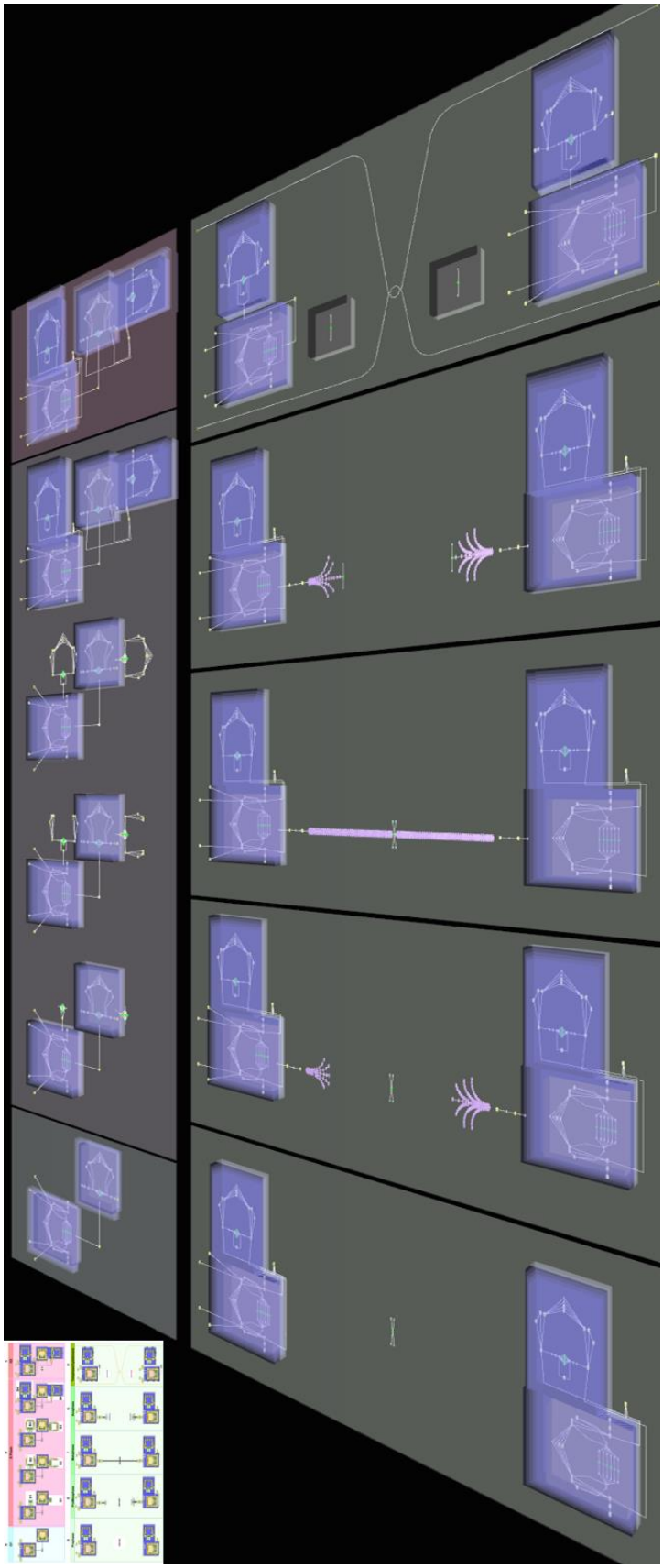


Figure 5.22: Interplay of 3D model of centrosomes. Upper left side represents the interplay of centrosome life cycle in 2D by yEd software. Centrosome structures depicted in 2D by mEPN scheme was visualised in 3D structures using BioLayout Express3D. For full-scale image, please use BioLayout Express3D to explore the pathway map file attached in CD.

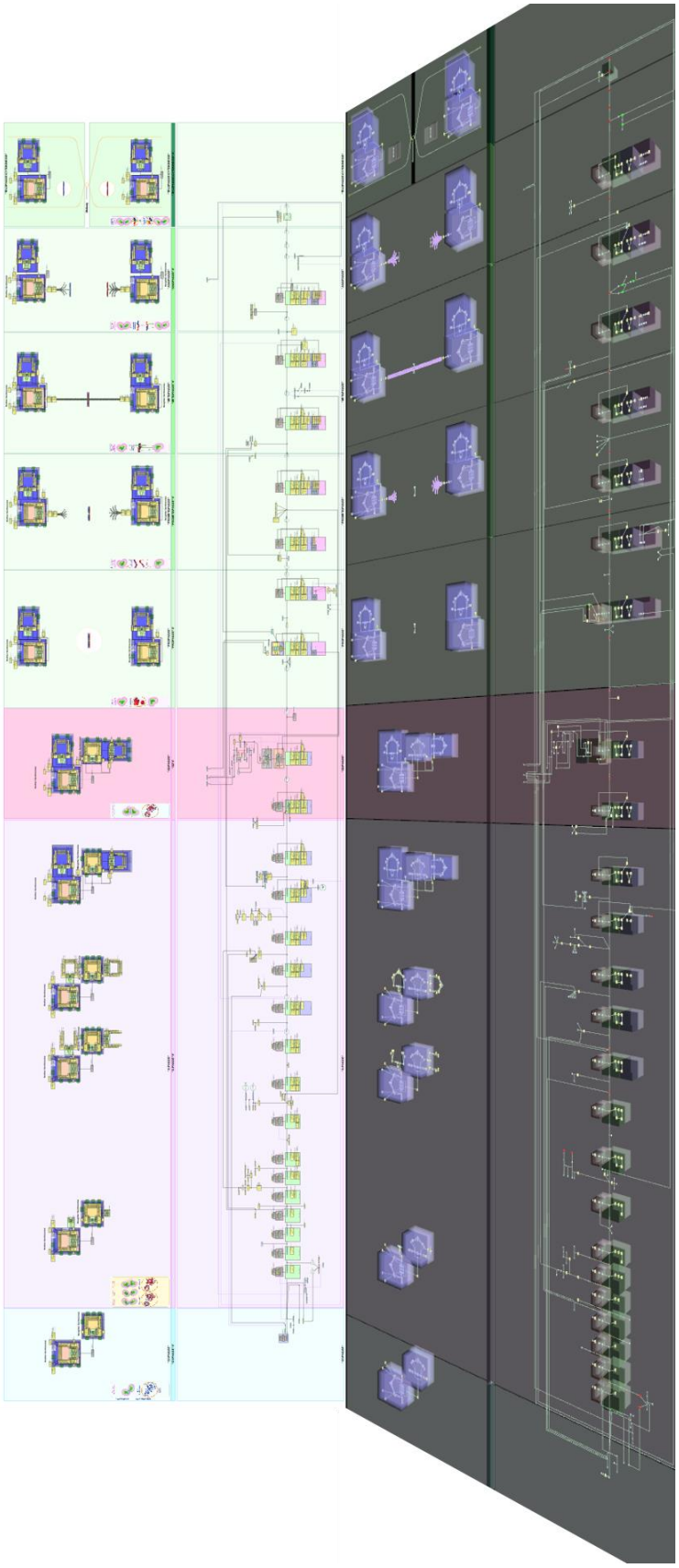


Figure 5.23: Pathway map of centrosome regulatory network demonstrated in 2D and 3D environment by yEd and BioLayout Express3D respectively. A) Regulatory network of centrosome-associated proteins demonstrated as cartoon diagrams and pathway maps by mEPN scheme. B) Translated '.graphmal' files of pathway maps and cartoon figures demonstrated in 3D environment with identical positions parallel with 2D images.

Discussion

Analysis of the localisation of unknown protein in Chapter 4 implicated that a number of them (e.g. FAM111B) may have a putative role in the centrosome life cycle. To examine their putative positions in the regulatory network of the centrosome-associated proteins, a large-scale pathway was developed including 117 important proteins.

The complex regulatory network of the protein-protein interactions involved in the centrosome life cycle is difficult to demonstrate accurately. The entire pathway map includes 117 proteins i.e. the development of procentrioles, recruitment of microtubule, and the redistribution of PCM between mother and matured daughter centrosomes.

This map gives a clear guide to the structure of centrosome and regulatory system that conduct the biological activities following the progression of cell cycle. For example, the disjunction of paired parental centrosomes is conducted by a complicated regulatory protein-protein interaction network (Figure 5.15B). However, the biological event of paired parental centrosome disjunction can be simply demonstrated by the red dash line in connection with parental centrosomes showing an ongoing process of disjunction. Therefore, demonstrating the centrosome-associated biological events as a pathway can enhance the information accessibility.

Construction of a large-scale regulatory network of centrosome-associated proteins in the cell cycle

The aim of constructing a large-scale biological process network is to create a pathway resource providing consensus view from literature curation. In order to continuously update the role of proteins in the centrosome life cycle, the pathway scheme needs to be designed in a way that optimisation and expansion of network structure can be made with updates by new publications and research observations. To support the work, the pathway software has to be user friendly, easy to access i.e. yEd.

To find the pathway model using approach that could work the needs of this thesis, a number of modelling schemes were evaluated e.g. SBGN and mEPN. The mEPN and SBGN schemes both provide similar network structure but different ways to visualise a biological processes. Similar concepts are used in the two systems to depict biological process, including components, processes, relationships and cellular compartments.

In comparison with SBGN scheme, a number of features in mEPN scheme make it suitable for this task. Firstly, for the interactions that are widely distributed in biological processes, each unique notation is only depicted once enabling efficient use of space in the pathway map. For example, PLK1 is involved in multiple reactions in this pathway map, including maturation of daughter centrioles (Figure 5.12), regulation of cohesin between parental centrosomes (Figure 5.13), disjunction of paired parental centrosomes (Figure 5.15), etc. Secondly, the use of colour and one or two letters to represent biological reactions in the pathway map provide enhanced visual accessibility. It is very useful to identify particular biological processes within the pathway, such as process nodes 'B' for binding with yellow colour and 'Ub' for ubiquitisation with light blue colour (Figure 5.2C). Thirdly, the compatibility of computational simulation in 3D aids in further analysis. Since there is a growing market of the computational graphic industries, such as gaming or animation, lots of technologies were developed to support the visualisations of data analysis in virtual 3D environments. A good case is that BioLayout *Express*^{3D} is that employs a 3D API (Application programming interface) OpenGL, providing powerful platform to analyse different types of 'omics' data visualised as a network structure [112,113].

Therefore, BioLayout *Express*^{3D} can import the pathway map saved as '.graphml' format and visualise the file in 2D and 3D. The translation of a pathway map from 2D into 3D environment allows the map to be rotated and zoomed in from multiple angles for particular biological events of interest (Figures 5.20, 5.21, 5.22, and 5.23). In addition the parent and children nodes connected to a node of interest can also be explored in a large-scale pathway network in 3D environment as well. This enhances the information accessibility in studying the regulatory logic of a protein or gene on interest (Figures 5.21 and 5.23).

In addition, the power of mEPN scheme is that it can be used to demonstrate not only the regulatory network but also the structure of centrosome at different stages in 2D and 3D, something not possible with SBGN diagrams (Figure 5.23). Owing to a number of advantages discussed above, the mEPN scheme was considered to satisfy the need of building a large-scale regulatory network of centrosome with the yEd software [177].

The genes associated with daughter centriole development are up-regulated; the regulatory genes that control the number of centrosomes are consistently expressed

In order to investigate the mechanisms that regulate centrosome biogenesis and duplication at transcriptional level, the list of genes regulated in array dataset (Chapter 2) and the

centrosome pathway map were summarised into a number of groups, according to expression pattern and the location of the protein in the centrosome highlighted on the pathway (Table 5.2 and Figure 5.20).

There are a total of 33 centrosomal genes, which are up-regulated during the S/G_2/M-phase duplication (Table 5.2). This includes the centrosome-associated genes involved in the early stage of centrosome development, such as *SASS6*, *CCP110*, *RAD21*, and *TUBG1*. *SASS6* is involved in the cartwheel structure development, *CCP110* is localised at the distal end of procentriole that inhibits the over extending centriole structure (Table 5.2) [344,345,381]. In addition, *RAD21* is localised at the cohesin complex and not only maintains the pairing of parental centrioles but also avoids the re-duplication of centrosome [270,414,415], and the protein *TUBG1* that belongs to γ-TuRC complexes [468].

For the proteins positioned in the inner and outer layer of PCM, many of their genes are also up-regulated in preparation to surround the procentriole such as *CEP120* (inner layer of PCM) [374,420] and *NEDD1* (outer layer of PCM) [358,361-363,469,470] involved in the organisation of PCM structure and recruitment of microtubule, respectively.

Of those genes, which have a cell cycle associated expression pattern as mentioned above, it suggests that the proteins are essential for the development of daughter centriole structure present only in S-phase. In contrast, consistent expression of these genes may potentially increase the possibility of centrosome re-duplication, therefore the nature of their expression behaviour only appear to be up-regulated across the G_1/S-border (Table 5.2).

For genes, which are constantly expressed according to the array dataset in this thesis (Table 5.2). It is interesting that 11 of 18 gens are involved in the regulation of centrosome duplication guaranteeing that the centrosomes only duplicate once per cell cycle (Table 5.2).

It appears that the genes, which are essential for the structural development in centrosome duplication, are up-regulated during the G_1/S- and S-phase.

A new module for centrosome duplication and biogenesis that improves the current understanding of centrosome biology

To see whether that this has been proposed previously, further literature review was undertaken to provide the current understanding of centrosome biogenesis and duplication. The majority of experimental approaches in the last 20 years are mostly aimed to identify the centrosome-associated proteins, suggesting that they are mainly involved in the studying at

structural level of the centrosome biogenesis, duplication, and the regulation of chromosome segregation in M-phase [52,337,471-474]. However, they are mainly focused on the proteomic level ignoring the mechanism that conducts the regulation between transcriptomic and proteomic level [52,337,471-474]. Therefore, little is known about the regulatory logic that governs the 'global scale networks of centrosome biology' covering the gene expression and protein-protein interactions networks.

To develop a model that can be used to explain centrosome biology, a methodology that combines the pathway mapping (i.e. a large-scale centrosome pathway map) and transcriptomic analysis (i.e. array experiments that provides time series transcriptomic profile) was first introduced in this thesis (Table 5.2 and Figure 5.20). It is interesting that the new model suggests a multiple level regulation between transcriptomic and proteomic activities controlling the number of centrosomes in different stage of the cell cycle.

Chapter 6.

Conclusions

Overview of major approaches

The work described in this thesis explores the use of systems-level analyses to identify and study a number of novel cell cycle-associated genes. Specifically, the work outlined in Chapter 2 described the identification of genes that may function in the cell cycle regulation. Identification of cell cycle-associated genes at a transcriptional level has been performed in human cells since 1999, where serum deprivation was used to generate synchronised human fibroblasts for array analysis. We were able to discover cell cycle-associated genes using microarray technology capable of investigating the expression profiles of all human genes in parallel. In order to provide samples for two microarray analyses, fibroblasts were selected. These cells were selected because they responded to serum-deprivation and arrested at a quiescent stage and then re-entered the cell cycle after serum refeeding with at least one synchronised cell cycle within 24 hours. In contrast, a transformed cell line THP-1 was found to be difficult to get to synchronise by serum stimulation. Furthermore, the cell cycle of THP-1 was longer than 24 hours, and it had been suggested that a prolonged period of one complete cell cycle was difficult to maintain. For HEK293T, the transfection with the SV40 virus forces cells to skip G_0-phase to enter the cell cycle even after serum deprivation suggesting that HEK293T was not an ideal cell type for this experiment either.

Known cell cycle-regulated genes were initially tested by RT-qPCR. Subsequently two repeatable array experiments were performed and analysed by network structure analysis in BioLayout *Express*^{3D} after quality assessment in the R environment. RMA normalisation minimised the technical variation between array samples. Animation of the network was used to demonstrate that the up-regulation of genes follows the order of a sequence of time points from G_0- (0 h) to M- phase (24 h). Using GO term enrichment analysis to profile the feature of clustered genes in four major groups, a higher proportion of genes placed in the S/G_2/M-phase cluster were associated with the 'cell cycle' term reflecting the re-entry of the cell cycle. Results of the network structure analysis identified four major clusters of genes including the G_0-, the early and late stage of G_1-, and S/G_2/M-phase. Our group worked on categorising a parts list of S/G_2/M-phase associated genes suggesting that 78, 69, 75, and 484 of a total of 706 genes are putative, unknown, unrelated, and known cell cycle-associated genes, respectively.

Chapter 3 describes the utilisation of the RTCA system in real time RNAi screening to determine the role of targeted genes in dividing cells. The RTCA system is a highly sensitive real time cell proliferation monitoring system. Electron sensors arrayed on the plates evaluate the number of cell population by detecting the electronic resistance from the lipid bilayer cell membrane. This allows the recording of the change in the cell population as CI declining was used to determine the effect of gene silencing on cell proliferation.

In order to find the best conditions for gene silencing with minimal cytotoxicity, that might potentially decreased the accuracy of RNAi screening, silenceMag™, a positively charged practice transfection reagent, was used to deliver dsRNA molecules into NHDF cells targeting the genes which were thought to be important in cell proliferation in Chapter 2. In contrast, liposome conducted transfection was found to arrest cell proliferation according to the profile of cell proliferation arrest monitored using the RTCA system. In addition, to avoid the off-target effects induced by siRNA mediated gene silencing, esiRNA was used to target the genes of interest. SYTOX® blue staining of cells that showed a higher cell viability was determined by flow cytometry.

A total of 42 unknown genes of interest were knocked down by RNAi, and verified by qPCR. 16 and 3 of them were classified as activators or regulators after knocking down, respectively. The hits on target genes in the RNAi screening were determined by Z-scoring cut-off. In addition, a time series of transformed CI over time series as CIGR and CIGR with respect to the negative control were used to demonstrate the time point at which there was a maximum effect on cell proliferation after transfection.

Protein localisation described in Chapter 4 was performed using HEK293T cells to determine their role in the cell proliferation. To provide an effective expression system, selected genes in chapter 3 were expressed in HEK293T cells using Gateway cloning technology. To demonstrate the control of proteins localisation, centrosome-regulated proteins e.g. CEP55, CEP57, CEP85, and CPM1 were expressed. The comparison between proteins of interest and control demonstrated similar localisation pattern implying their potential involvement in centrosome regulation e.g. FAM111B.

In order to conclude the discovery of the genes that may potentially have a putative role in centrosome regulation, a large-scale pathway network was constructed. A total of 117 proteins known to be important to the centrosome were included in the pathway. The pathway included the regulation of centrosome development including the recruitment of

proteins for procentriole development, redistribution of PCM during the procentrioles maturation, and disjunction of paired parental centrosomes serving for duplicated sister chromatids segregation across the entire cell cycle.

Chapter 5 also summarises the regulation network of the centrosome pathway map at a transcriptional level as determined by the microarray dataset. It is known that to avoid the re-duplication of daughter centrioles, which may lead to unequal segregation of duplicated centrosomes, it is important to maintain the pairing of parental centrosomes by regulation at the transcriptional level. To avoid the over development of procentrioles, the genes encoding proteins, which are involved in the development of daughter centriole structure, are periodically expressed. In contrast, for proteins which are involved in the inhibition of daughter centriole re-duplication, the majority of their genes are consistently expressed. This concept of a regulatory mechanism has not been suggested before.

To further testify the originality of this hypothesis, literature review was performed. However, the majority of experimental approaches are mainly focused on the structure level of centrosome biology. In contrast, the strategy that uses the pathway map to investigate the regulation mechanism at transcriptional level here is first used in this thesis.

In summary, two genes of interest (i.e. *FAM111B* and *KIAA1549L*) have been identified that they may have putative roles in the centrosome biology using array analysis, RNAi screening, and a large-scale pathway map. In addition, a new module that improves the understanding of centrosome biology was first discovered showing that the mechanism which controls the centrosome duplication may be involved in multiple levels of gene expression behaviour. Literature review suggests that this a new methodology first used here.

Challenges

The challenge in this project has been to use different methodologies including cell cycle synchronisation follower by array analysis, RNAi screening, and pathway mapping to investigate the cell cycle associated transcription.

In order to provide the samples for transcriptomic analysis, in which the cellular regulation at transcriptional level reflects the nature of cell physiology, the first challenge was to choose non-chemical dependent techniques as it has been suggested that the off-target effects of chemical components may induce non-specific biological events. Of those non-chemical techniques, only serum deprivation can provide enough quantity of sample. In addition, the

use of primary cells line can profile the gene expression periodicities as it has been suggested that the cell cycle-associated genes in transformed cell lines are mostly over expressed.

There are two array experiments in this work, where the time series datasets are at different intervals. It is difficult to adopt conventional statistical analysis identifying the gene expression without enough number of repeats. In doing so, we introduce the network based analysis using BioLayout *Express*^{3D}. This allows us to cluster the gene of interest.

However, the complexity of various annotations highlighted the difficulty in integrating the information, e.g. UniPort and ENSEMBL ID. Therefore, one of the biggest challenges is to systematically categorise the gene clusters. To provide a solution, Freeman's group equally share the work of manual curation and annotation of a large number of clustered genes finalizing into a list of 706 genes in association with the cell biology during S/G_2/M-phase.

Of these 706 genes, 42 of the most interesting were selected for further analysis as literature mining suggest that were not previously associated with in the cell cycle. To test whether they are important in the cell proliferation, the RNAi screening was used to see the loss of function effect on cell division using RTCA system. However, a number of elements may affect the results of RNAi screening e.g. the liposome dependent transfection that arrests the cell proliferation and the off-target effects induced by siRNA components. To minimize the artifacts, the solution is to use magnet-assisted transfection delivering esiRNA components avoiding cytotoxicity of liposome-based reagents and to dilute the off-target effects of siRNA components. The result of RNAi screening identified 19 genes which appear to play important role in cell proliferation.

To further investigate the genes identified by RNAi, their protein products were localized using GFP tags in the cells. In this thesis, fibroblast cells were used for the array analysis and RNAi screening. However, transfection efficiencies in fibroblast cells were poor. In contrast, the subcellular protein localisation in HEK293T cells shows robust transfection efficiency. Therefore, the HEK293T were chosen for the experiments. The result of subcellular protein localisation studies suggests that the protein products of *FAM111B* and *KIAA1549L* are localised to the centrosome.

Encouraged by this approach, literature mining of centrosome-associated proteins was carried out predicting their putative roles in the centrosome biology. However, the majority of literature were mostly emphasised on one or two single pathway of a specific protein

relevant to centrosome biology. To provide integrated information of centrosome biology from global aspects, a large-scale pathway model was developed including the protein-protein networks that comprise of centrosome biogenesis, duplication, and the mitotic activities.

As the pathway is highly complicated, it was a challenge to illustrate the biological events at system level where the network composed of components (i.e. proteins) and edges (i.e. direction of biological events). Therefore, the best solution was to draw a flowchart at a structure level, where the change in centrosome structure at different stages. It is interesting that the centrosome biology at transcriptional level can be summarised into different patterns according to their localisation in the centrosome structure. The majority of regulators that control centrosome numbers are consistently expressed. In contrast, the genes that periodically expressed mostly play essential roles in daughter centriole duplication e.g. gene of the cartwheel structure-associated proteins are up-regulated in S/G_2/M-phase.

Future work

In future plan, we would like to screen the rest of cell cycle clustered genes identified in our array dataset, identifying more putative genes that may be involved in cell division. Genes that appears to be involved in cell division will be co-localised with the known cell cycle-associated protein in the cell cycle, e.g. gamma-tubulin.

Encouraged by the success of using a pathway map we are able to better understand the regulatory mechanisms governing the centrosome-associated protein at transcriptional level, a full scale protein-protein interaction pathway that covers the cell cycle will be developed in combination with our array dataset to investigate the mechanisms that drive the cascade of cell cycle events. The computational simulation will be carried out to predict the regulation behaviour of protein of interest according to the result of RNAi screening and subcellular protein localisation.

Overall, each of the chapters was organised following a logical flow in providing evidence to further investigate the putative function of novel genes associated with cell proliferation. The centrosome regulatory network analysis provided a global scale for interpreting centrosome regulation at a transcriptional level. Further evidence and pathway modelling of putative novel cell cycle-associated genes is needed to fill in the last pieces of information to determine their role in the regulatory network of centrosomes. Eventually the hope is that

such models will be able to provide predictive power to identify protein function based on further characterisation of these proteins of previously unknown function.

References

1. Schleiden MJ (1838) Archiv für Anatomie, Physiologie und Wissenschaftliche Medicin.
2. Schwann T (1839) Mikroskopische Untersuchungen über die Übereinstimmung in der Struktur und dem Wachstum der Tiere und Pflanzen. Berlin.
3. Mazzarello P (1999) A unifying concept: the history of cell theory. Nat Cell Biol 1: E13-15.
4. Mayr E (1982) The Growth of the Biological Thought. Cambridge.
5. Dubrovsky JG, Ivanov VB (2003) Celebrating 50 years of the cell cycle. Nature 426: 759.
6. Howard A (1951) Nuclear incorporation of P32 as demonstrated by autoradiographs. Experimental Cell Research 2: 178–187.
7. Jackman J, O'Connor PM (2001) Methods for synchronizing cells at specific stages of the cell cycle. Curr Protoc Cell Biol Chapter 8: Unit 8 3.
8. Ruffy MB, Kunnavatana SS, Koch RJ (2006) Effects of tamoxifen on normal human dermal fibroblasts. Arch Facial Plast Surg 8: 329-332.
9. Tsuchiya S, Yamabe M, Yamaguchi Y, Kobayashi Y, Konno T, et al. (1980) Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). Int J Cancer 26: 171-176.
10. Bernard S, Herzel H (2006) Why do cells cycle with a 24 hour period? Genome Inform 17: 72-79.
11. Biosciences B (2015) The Cell Cycle. In: Cycle TC, editor: BD Biosciences.
12. Pardee AB (1974) A restriction point for control of normal animal cell proliferation. Proc Natl Acad Sci U S A 71: 1286-1290.
13. Morgan DO (2007) The cell cycle : principles of control. London

Sunderland, MA: Published by New Science Press in association with Oxford University Press ;

Distributed inside North America by Sinauer Associates, Publishers. xxvii, 297 p. p.
14. Cimprich KA, Cortez D (2008) ATR: an essential regulator of genome integrity. Nat Rev Mol Cell Biol 9: 616-627.
15. Forsburg SL (2008) The MCM helicase: linking checkpoints to the replication fork. Biochem Soc Trans 36: 114-119.
16. Sudo T, Nitta M, Saya H, Ueno NT (2004) Dependence of paclitaxel sensitivity on a functional spindle assembly checkpoint. Cancer Res 64: 2502-2508.
17. Cleveland DW, Mao Y, Sullivan KF (2003) Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. Cell 112: 407-421.
18. Hunt T (2004) The discovery of cyclin (I). Cell 116: S63-64, 61 p following S65.
19. Evans T (2004) The discovery of cyclin (II). Cell 116: S65, 61 p following S65.
20. Evans T, Rosenthal ET, Youngblom J, Distel D, Hunt T (1983) Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. Cell 33: 389-396.
21. Malumbres M, Barbacid M (2009) Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev Cancer 9: 153-166.
22. Coller HA, Sang L, Roberts JM (2006) A new description of cellular quiescence. PLoS Biol 4: e83.
23. Roussel MF (1999) The INK4 family of cell cycle inhibitors in cancer. Oncogene 18: 5311-5317.
24. Denicourt C, Dowdy SF (2004) Cip/Kip proteins: more than just CDKs inhibitors. Genes Dev 18: 851-855.
25. Krause A, Hoffmann I (2010) Polo-like kinase 2-dependent phosphorylation of NPM/B23 on serine 4 triggers centriole duplication. PLoS One 5: e9849.

26. Coverley D, Laman H, Laskey RA (2002) Distinct roles for cyclins E and A during DNA replication complex assembly and activation. Nat Cell Biol 4: 523-528.
27. Draetta G, Luca F, Westendorf J, Brizuela L, Ruderman J, et al. (1989) Cdc2 protein kinase is complexed with both cyclin A and B: evidence for proteolytic inactivation of MPF. Cell 56: 829-838.
28. Ewen ME (2000) Where the cell cycle and histones meet. Genes Dev 14: 2265-2270.
29. Hinchcliffe EH, Li C, Thompson EA, Maller JL, Sluder G (1999) Requirement of Cdk2-cyclin E activity for repeated centrosome reproduction in Xenopus egg extracts. Science 283: 851-854.
30. Fisk HA, Winey M (2001) The mouse Mps1p-like kinase regulates centrosome duplication. Cell 106: 95-104.
31. Mott ML, Berger JM (2007) DNA replication initiation: mechanisms and regulation in bacteria. Nat Rev Microbiol 5: 343-354.
32. Bell SP, Dutta A (2002) DNA replication in eukaryotic cells. Annu Rev Biochem 71: 333-374.
33. Arias EE, Walter JC (2007) Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. Genes Dev 21: 497-518.
34. Fletcher RJ, Bishop BE, Leon RP, Sclafani RA, Ogata CM, et al. (2003) The structure and function of MCM from archaeal M. Thermoautotrophicum. Nat Struct Biol 10: 160-167.
35. Mehta GD, Kumar R, Srivastava S, Ghosh SK (2013) Cohesin: functions beyond sister chromatid cohesion. FEBS Lett 587: 2299-2312.
36. Tsou MF, Stearns T (2006) Mechanism limiting centrosome duplication to once per cell cycle. Nature 442: 947-951.
37. Bornens M (2012) The centrosome in cells and organisms. Science 335: 422-426.
38. Bettencourt-Dias M, Glover DM (2009) SnapShot: centriole biogenesis. Cell 136: 188-188 e181.
39. Lacey KR, Jackson PK, Stearns T (1999) Cyclin-dependent kinase control of centrosome duplication. Proc Natl Acad Sci U S A 96: 2817-2822.
40. Matsumoto Y, Hayashi K, Nishida E (1999) Cyclin-dependent kinase 2 (Cdk2) is required for centrosome duplication in mammalian cells. Curr Biol 9: 429-432.
41. Meraldi P, Lukas J, Fry AM, Bartek J, Nigg EA (1999) Centrosome duplication in mammalian somatic cells requires E2F and Cdk2-cyclin A. Nat Cell Biol 1: 88-93.
42. Charrier-Savournin FB, Chateau MT, Gire V, Sedivy J, Piette J, et al. (2004) p21-Mediated nuclear retention of cyclin B1-Cdk1 in response to genotoxic stress. Mol Biol Cell 15: 3965-3976.
43. Bastos RN, Barr FA (2010) Plk1 negatively regulates Cep55 recruitment to the midbody to ensure orderly abscission. J Cell Biol 191: 751-760.
44. Fabbro M, Zhou BB, Takahashi M, Sarcevic B, Lal P, et al. (2005) Cdk1/Erk2- and Plk1-dependent phosphorylation of a centrosome protein, Cep55, is required for its recruitment to midbody and cytokinesis. Dev Cell 9: 477-488.
45. Chang YC, Wu CH, Yen TC, Ouyang P (2012) Centrosomal protein 55 (Cep55) stability is negatively regulated by p53 protein through Polo-like kinase 1 (Plk1). J Biol Chem 287: 4376-4385.
46. Garbe C (2012) Patterns of sun protection for young children: do we deliver the right sun-protection messages? Br J Dermatol 166: 710.
47. Nigg EA (2002) Centrosome aberrations: cause or consequence of cancer progression? Nat Rev Cancer 2: 815-825.
48. Sluder G, Nordberg JJ (2004) The good, the bad and the ugly: the practical consequences of centrosome amplification. Curr Opin Cell Biol 16: 49-54.

49. Kuriyama R, Borisy GG (1981) Centriole cycle in Chinese hamster ovary cells as determined by whole-mount electron microscopy. J Cell Biol 91: 814-821.
50. Tanos BE, Yang HJ, Soni R, Wang WJ, Macaluso FP, et al. (2013) Centriole distal appendages promote membrane docking, leading to cilia initiation. Genes Dev 27: 163-168.
51. Kodani A, Salome Sirerol-Piquer M, Seol A, Garcia-Verdugo JM, Reiter JF (2013) Kif3a interacts with Dynactin subunit p150 Glued to organize centriole subdistal appendages. EMBO J 32: 597-607.
52. Pihan GA (2013) Centrosome dysfunction contributes to chromosome instability, chromoanagenesis, and genome reprograming in cancer. Front Oncol 3: 277.
53. Hudson DF, Marshall KM, Earnshaw WC (2009) Condensin: Architect of mitotic chromosomes. Chromosome Res 17: 131-144.
54. Kimura K, Hirano M, Kobayashi R, Hirano T (1998) Phosphorylation and activation of 13S condensin by Cdc2 in vitro. Science 282: 487-490.
55. KITAGAWA D (2015) Cell cycle-dependent centriole formation.
56. Fukasawa K (2007) Oncogenes and tumour suppressors take on centrosomes. Nat Rev Cancer 7: 911-924.
57. Azimzadeh J, Marshall WF (2010) Building the centriole. Curr Biol 20: R816-825.
58. Maiato H, DeLuca J, Salmon ED, Earnshaw WC (2004) The dynamic kinetochore-microtubule interface. J Cell Sci 117: 5461-5477.
59. Holland AJ, Taylor SS (2006) Cyclin-B1-mediated inhibition of excess separase is required for timely chromosome disjunction. J Cell Sci 119: 3325-3336.
60. Holt LJ, Krutchinsky AN, Morgan DO (2008) Positive feedback sharpens the anaphase switch. Nature 454: 353-357.
61. Hetzer MW (2010) The nuclear envelope. Cold Spring Harb Perspect Biol 2: a000539.
62. Berridge MJ (2014) Cell Signalling Biology. Cell Signalling Biology 9: 45.
63. Carlton JG, Martin-Serrano J (2007) Parallels between cytokinesis and retroviral budding: a role for the ESCRT machinery. Science 316: 1908-1912.
64. Lee HH, Elia N, Ghirlando R, Lippincott-Schwartz J, Hurley JH (2008) Midbody targeting of the ESCRT machinery by a noncanonical coiled coil in CEP55. Science 322: 576-580.
65. Maskos U, Southern EM (1992) Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesised in situ. Nucleic Acids Res 20: 1679-1684.
66. Bumgarner R (2013) Overview of DNA microarrays: types, applications, and their future. Curr Protoc Mol Biol Chapter 22: Unit 22 21.
67. Crampton J, Humphries S, Woods D, Williamson R (1980) The isolation of cloned cDNA sequences which are differentially expressed in human lymphocytes and fibroblasts. Nucleic Acids Res 8: 6007-6017.
68. Miller JK, Barnes WM (1986) Colony probing as an alternative to standard sequencing as a means of direct analysis of chromosomal DNA to determine the spectrum of single-base changes in regions of known sequence. Proc Natl Acad Sci U S A 83: 1026-1030.
69. Gergen JP, Stern RH, Wensink PC (1979) Filter replicas and permanent collections of recombinant DNA plasmids. Nucleic Acids Res 7: 2115-2136.
70. Craig AG, Nizetic D, Hoheisel JD, Zehetner G, Lehrach H (1990) Ordering of cosmid clones covering the herpes simplex virus type I (HSV-I) genome: a test case for fingerprinting by hybridisation. Nucleic Acids Res 18: 2653-2660.
71. DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, et al. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 14: 457-460.

72. (2015) Array Production. In: robot mp, editor: The scripps reseacrh institute.
73. Merrifield B (1986) Solid phase synthesis. Science 232: 341-347.
74. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, et al. (1991) Light-directed, spatially addressable parallel chemical synthesis. Science 251: 767-773.
75. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ (1999) High density synthetic oligonucleotide arrays. Nat Genet 21: 20-24.
76. Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, et al. (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci U S A 91: 5022-5026.
77. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, et al. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 14: 1675-1680.
78. Nuwaysir EF, Huang W, Albert TJ, Singh J, Nuwaysir K, et al. (2002) Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. Genome Res 12: 1749-1755.
79. Singh-Gasson S, Green RD, Yue Y, Nelson C, Blattner F, et al. (1999) Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. Nat Biotechnol 17: 974-978.
80. Blanchard AP, Kaiser RJ, Hood LE (1996) High-density oligonucleotide arrays. Biosensors and Bioelectronics 11: 687-690.
81. Technology AM (2015) In Situ Synthesis Printing Process.
82. Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, et al. (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol 19: 342-347.
83. Ferguson JA, Steemers FJ, Walt DR (2000) High-density fiber-optic DNA random microsphere array. Anal Chem 72: 5618-5624.
84. Michael KL, Taylor LC, Schultz SL, Walt DR (1998) Randomly ordered addressable high-density optical sensor arrays. Anal Chem 70: 1242-1248.
85. Steemers FJ, Ferguson JA, Walt DR (2000) Screening unlabeled DNA targets with randomly ordered fiber-optic gene arrays. Nat Biotechnol 18: 91-94.
86. Walt DR (2000) Techview: molecular biology. Bead-based fiber-optic arrays. Science 287: 451-452.
87. Walt DR (2006) Fiber optic array biosensors. Biotechniques 41: 529, 531, 533 passim.
88. Gunderson KL, Kruglyak S, Graige MS, Garcia F, Kermani BG, et al. (2004) Decoding randomly ordered DNA arrays. Genome Res 14: 870-877.
89. Davis PK, Ho A, Dowdy SF (2001) Biological methods for cell-cycle synchronization of mammalian cells. Biotechniques 30: 1322-1326, 1328, 1330-1321.
90. Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, et al. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell 2: 65-73.
91. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, et al. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol Biol Cell 9: 3273-3297.
92. Rustici G, Mata J, Kivinen K, Lio P, Penkett CJ, et al. (2004) Periodic gene expression program of the fission yeast cell cycle. Nat Genet 36: 809-817.
93. Laub MT, McAdams HH, Feldblyum T, Fraser CM, Shapiro L (2000) Global analysis of the genetic network controlling a bacterial cell cycle. Science 290: 2144-2148.
94. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, et al. (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 13: 1977-2000.

95. Ishida S, Huang E, Zuzan H, Spang R, Leone G, et al. (2001) Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis. Mol Cell Biol 21: 4684-4699.
96. Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, et al. (1999) The transcriptional program in the response of human fibroblasts to serum. Science 283: 83-87.
97. Cho RJ, Huang M, Campbell MJ, Dong H, Steinmetz L, et al. (2001) Transcriptional regulation and function during the human cell cycle. Nat Genet 27: 48-54.
98. Bar-Joseph Z, Siegfried Z, Brandeis M, Brors B, Lu Y, et al. (2008) Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. Proc Natl Acad Sci U S A 105: 955-960.
99. Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467-470.
100. Barnhart BJ (1989) The Department of Energy (DOE) Human Genome Initiative. Genomics 5: 657-660.
101. Watson JD, Jordan E (1989) The Human Genome Program at the National Institutes of Health. Genomics 5: 654-656.
102. Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, et al. (2004) Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS Biol 2: E7.
103. Deonarine K, Panelli MC, Stashower ME, Jin P, Smith K, et al. (2007) Gene expression profiling of cutaneous wound healing. J Transl Med 5: 11.
104. Do JH, Choi DK (2006) Normalization of microarray data: single-labeled and dual-labeled arrays. Mol Cells 22: 254-261.
105. Irizarry RA, Bolstad BM (2003) Summaries of Affymetrix GeneChip probe level data. 31: e15.
106. Irizarry RA, Ooi SL, Wu Z, Boeke JD (2003) Use of mixture models in a microarray-based screening procedure for detecting differentially represented yeast mutants. Stat Appl Genet Mol Biol 2: Article1.
107. Troyanskaya OG, Garber ME, Brown PO, Botstein D, Altman RB (2002) Nonparametric methods for identifying differentially expressed genes in microarray data. Bioinformatics 18: 1454-1461.
108. Mutch DM, Berger A, Mansourian R, Rytz A, Roberts MA (2001) Microarray data analysis: a practical approach for selecting differentially expressed genes. Genome Biol 2: PREPRINT0009.
109. Selvaraj S, Natarajan J (2011) Microarray data analysis and mining tools. Bioinformation 6: 95-99.
110. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95: 14863-14868.
111. Kintigh KW, Ammerman AJ (1982) Heuristic Approaches to Spatial-Analysis in Archaeology. American Antiquity 47: 31-63.
112. Theocharidis A, van Dongen S, Enright AJ, Freeman TC (2009) Network visualization and analysis of gene expression data using BioLayout Express(3D). Nat Protoc 4: 1535-1550.
113. Freeman TC, Goldovsky L, Brosch M, van Dongen S, Maziere P, et al. (2007) Construction, visualisation, and clustering of transcription networks from microarray expression data. PLoS Comput Biol 3: 2032-2042.
114. Dongen Sv (2000) Graph clustering by flow simulation. Netherlands: Centre for Mathematics and Computer Science.
115. Zhang JD, Koerner C, Bechtel S, Bender C, Keklikoglou I, et al. (2011) Time-Resolved Human Kinome RNAi Screen Identifies a Network Regulating Mitotic-Events as Early Regulators of Cell Proliferation. PLoS One 6.

116. Napoli C, Lemieux C, Jorgensen R (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell 2: 279-289.
117. Guo S, Kemphues KJ (1995) par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell 81: 611-620.
118. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, et al. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391: 806-811.
119. Hamilton AJ, Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286: 950-952.
120. Zamore PD, Tuschl T, Sharp PA, Bartel DP (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell 101: 25-33.
121. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, et al. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411: 494-498.
122. Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409: 363-366.
123. Chu CY, Rana TM (2006) Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. PLoS Biol 4: e210.
124. Lord CJ, Martin SA, Ashworth A (2009) RNA interference screening demystified. J Clin Pathol 62: 195-200.
125. Echeverri CJ, Perrimon N (2006) High-throughput RNAi screening in cultured cells: a user's guide. Nat Rev Genet 7: 373-384.
126. Mohr S, Bakal C, Perrimon N (2010) Genomic screening with RNAi: results and challenges. Annu Rev Biochem 79: 37-64.
127. Zhang JD, Koerner C, Bechtel S, Bender C, Keklikoglou I, et al. (2011) Time-Resolved Human Kinome RNAi Screen Identifies a Network Regulating Mitotic-Events as Early Regulators of Cell Proliferation. PLoS ONE 6: e22176.
128. Iorns E, Lord CJ, Turner N, Ashworth A (2007) Utilizing RNA interference to enhance cancer drug discovery. Nat Rev Drug Discov 6: 556-568.
129. Sharma S, Rao A (2009) RNAi screening: tips and techniques. Nat Immunol 10: 799-804.
130. Boutros M, Kiger AA, Armknecht S, Kerr K, Hild M, et al. (2004) Genome-wide RNAi analysis of growth and viability in Drosophila cells. Science 303: 832-835.
131. Friedman A, Perrimon N (2006) A functional RNAi screen for regulators of receptor tyrosine kinase and ERK signalling. Nature 444: 230-234.
132. Kittler R, Pelletier L, Heninger AK, Slabicki M, Theis M, et al. (2007) Genome-scale RNAi profiling of cell division in human tissue culture cells. Nat Cell Biol 9: 1401-1412.
133. Moffat J, Grueneberg DA, Yang X, Kim SY, Kloepfer AM, et al. (2006) A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell 124: 1283-1298.
134. Root DE, Hacohen N, Hahn WC, Lander ES, Sabatini DM (2006) Genome-scale loss-of-function screening with a lentiviral RNAi library. Nat Methods 3: 715-719.
135. Simmer F, Moorman C, van der Linden AM, Kuijk E, van den Berghe PV, et al. (2003) Genome-wide RNAi of C. elegans using the hypersensitive rrf-3 strain reveals novel gene functions. PLoS Biol 1: E12.
136. Sledz CA, Holko M, de Veer MJ, Silverman RH, Williams BR (2003) Activation of the interferon system by short-interfering RNAs. Nat Cell Biol 5: 834-839.

137. Gaynor JW, Campbell BJ, Cosstick R (2010) RNA interference: a chemist's perspective. Chem Soc Rev 39: 4169-4184.
138. Baulcombe D (2004) RNA silencing in plants. Nature 431: 356-363.
139. Elbashir SM, Harborth J, Weber K, Tuschl T (2002) Analysis of gene function in somatic mammalian cells using small interfering RNAs. Methods 26: 199-213.
140. Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, et al. (2003) Expression profiling reveals off-target gene regulation by RNAi. Nat Biotechnol 21: 635-637.
141. Semizarov D, Frost L, Sarthy A, Kroeger P, Halbert DN, et al. (2003) Specificity of short interfering RNA determined through gene expression signatures. Proc Natl Acad Sci U S A 100: 6347-6352.
142. Scacheri PC, Rozenblatt-Rosen O, Caplen NJ, Wolfsberg TG, Umayam L, et al. (2004) Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. Proc Natl Acad Sci U S A 101: 1892-1897.
143. Birmingham A, Anderson EM, Reynolds A, Ilsley-Tyree D, Leake D, et al. (2006) 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. Nat Methods 3: 199-204.
144. Buchholz F, Kittler R, Slabicki M, Theis M (2006) Enzymatically prepared RNAi libraries. Nat Methods 3: 696-700.
145. Atienza JM, Zhu J, Wang X, Xu X, Abassi Y (2005) Dynamic monitoring of cell adhesion and spreading on microelectronic sensor arrays. J Biomol Screen 10: 795-805.
146. Kittler R, Putz G, Pelletier L, Poser I, Heninger AK, et al. (2004) An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. Nature 432: 1036-1040.
147. Juul N, Szallasi Z, Eklund AC, Li Q, Burrell RA, et al. (2010) Assessment of an RNA interference screen-derived mitotic and ceramide pathway metagene as a predictor of response to neoadjuvant paclitaxel for primary triple-negative breast cancer: a retrospective analysis of five clinical trials. Lancet Oncol 11: 358-365.
148. Nir O, Bakal C, Perrimon N, Berger B (2010) Inference of RhoGAP/GTPase regulation using single-cell morphological data from a combinatorial RNAi screen. Genome Res 20: 372-380.
149. Zhang JH, Chung TD, Oldenburg KR (1999) A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. J Biomol Screen 4: 67-73.
150. Brideau C, Gunter B, Pikounis B, Liaw A (2003) Improved statistical methods for hit selection in high-throughput screening. J Biomol Screen 8: 634-647.
151. Neumann B, Walter T, Heriche JK, Bulkescher J, Erfle H, et al. (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature 464: 721-727.
152. Neumann B, Held M, Liebel U, Erfle H, Rogers P, et al. (2006) High-throughput RNAi screening by time-lapse imaging of live human cells. Nat Methods 3: 385-390.
153. Hutchins JR, Toyoda Y, Hegemann B, Poser I, Heriche JK, et al. (2010) Systematic analysis of human protein complexes identifies chromosome segregation proteins. Science 328: 593-599.
154. Sahin O, Frohlich H, Lobke C, Korf U, Burmester S, et al. (2009) Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. BMC Syst Biol 3: 1.
155. Quereda JJ, Martinez-Alarcon L, Mendoca L, Majado MJ, Herrero-Medrano JM, et al. (2010) Validation of xCELLigence real-time cell analyzer to assess compatibility in xenotransplantation with pig-to-baboon model. Transplant Proc 42: 3239-3243.

156. Slanina H, Konig A, Claus H, Frosch M, Schubert-Unkmeir A (2011) Real-time impedance analysis of host cell response to meningococcal infection. J Microbiol Methods 84: 101-108.
157. Smout MJ, Kotze AC, McCarthy JS, Loukas A (2010) A novel high throughput assay for anthelmintic drug screening and resistance diagnosis by real-time monitoring of parasite motility. PLoS Negl Trop Dis 4: e885.
158. Tsien RY (1998) The green fluorescent protein. Annu Rev Biochem 67: 509-544.
159. von Arnim AG, Deng XW, Stacey MG (1998) Cloning vectors for the expression of green fluorescent protein fusion proteins in transgenic plants. Gene 221: 35-43.
160. Grebenok RJ, Pierson E, Lambert GM, Gong FC, Afonso CL, et al. (1997) Green-fluorescent protein fusions for efficient characterization of nuclear targeting. Plant J 11: 573-586.
161. Bordonne R (2000) Functional characterization of nuclear localization signals in yeast Sm proteins. Mol Cell Biol 20: 7943-7954.
162. Minopoli G, de Candia P, Bonetti A, Faraonio R, Zambrano N, et al. (2001) The beta-amyloid precursor protein functions as a cytosolic anchoring site that prevents Fe65 nuclear translocation. J Biol Chem 276: 6545-6550.
163. Sheng Z, Lewis JA, Chirico WJ (2004) Nuclear and nucleolar localization of 18-kDa fibroblast growth factor-2 is controlled by C-terminal signals. J Biol Chem 279: 40153-40160.
164. Hori Y (2007) [Crystal structure of the Aequorea victoria green fluorescent protein]. Tanpakushitsu Kakusan Koso 52: 1768-1769.
165. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994) Green fluorescent protein as a marker for gene expression. Science 263: 802-805.
166. Ross-Macdonald P, Coelho PS, Roemer T, Agarwal S, Kumar A, et al. (1999) Large-scale analysis of the yeast genome by transposon tagging and gene disruption. Nature 402: 413-418.
167. Matsuyama A, Arai R, Yashiroda Y, Shirai A, Kamata A, et al. (2006) ORFeome cloning and global analysis of protein localization in the fission yeast Schizosaccharomyces pombe. Nat Biotechnol 24: 841-847.
168. Brasch MA, Hartley JL, Vidal M (2004) ORFeome cloning and systems biology: standardized mass production of the parts from the parts-list. Genome Res 14: 2001-2009.
169. Ohara O (2009) ORFeome cloning. Methods Mol Biol 577: 3-9.
170. Institute D-FC (2015) hORFeome Database.
171. Russell PJ (2006) iGenetics : a molecular approach. 2nd ed. San Francisco: Pearson/Benjamin Cummings,. pp. xxi, 842 p. ill. (chiefly col.) 829 cm. + 841 CD-ROM (844 843/844 in.).
172. Landy A (1989) Dynamic, structural, and regulatory aspects of lambda site-specific recombination. Annu Rev Biochem 58: 913-949.
173. Hartley JL, Temple GF, Brasch MA (2000) DNA cloning using in vitro site-specific recombination. Genome Res 10: 1788-1795.
174. Raza S, Robertson KA, Lacaze PA, Page D, Enright AJ, et al. (2008) A logic-based diagram of signalling pathways central to macrophage activation. BMC Syst Biol 2: 36.
175. Kohn KW (1999) Molecular interaction map of the mammalian cell cycle control and DNA repair systems. Mol Biol Cell 10: 2703-2734.
176. Kohn KW, Aladjem MI, Weinstein JN, Pommier Y (2006) Molecular interaction maps of bioregulatory networks: a general rubric for systems biology. Mol Biol Cell 17: 1-13.

177. Freeman TC, Raza S, Theocharidis A, Ghazal P (2010) The mEPN scheme: an intuitive and flexible graphical system for rendering biological pathways. BMC Syst Biol 4: 65.
178. Kitano H, Funahashi A, Matsuoka Y, Oda K (2005) Using process diagrams for the graphical representation of biological networks. Nat Biotechnol 23: 961-966.
179. Le Novere N, Hucka M, Mi H, Moodie S, Schreiber F, et al. (2009) The Systems Biology Graphical Notation. Nat Biotechnol 27: 735-741.
180. Mardin BR, Schiebel E (2012) Breaking the ties that bind: new advances in centrosome biology. J Cell Biol 197: 11-18.
181. Jaspersen SL, Stearns T (2008) Exploring the pole: an EMBO conference on centrosomes and spindle pole bodies. Nat Cell Biol 10: 1375-1378.
182. Crasta K, Surana U (2006) Disjunction of conjoined twins: Cdk1, Cdh1 and separation of centrosomes. Cell Div 1: 12.
183. Bader GD, Cary MP, Sander C (2006) Pathguide: a pathway resource list. Nucleic Acids Res 34: D504-506.
184. Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, et al. (2008) WikiPathways: pathway editing for the people. PLoS Biol 6: e184.
185. Vastrik I, D'Eustachio P, Schmidt E, Gopinath G, Croft D, et al. (2007) Reactome: a knowledge base of biologic pathways and processes. Genome Biol 8: R39.
186. Ogata H, Goto S, Fujibuchi W, Kanehisa M (1998) Computation with the KEGG pathway database. Biosystems 47: 119-128.
187. Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, et al. (2012) WikiPathways: building research communities on biological pathways. Nucleic Acids Res 40: D1301-1307.
188. Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, et al. (2009) Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res 37: D619-622.
189. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27-30.
190. Gormley P, Li K, Irwin GW (2007) Modelling molecular interaction pathways using a two-stage identification algorithm. Syst Synth Biol 1: 145-160.
191. Ruths D, Muller M, Tseng JT, Nakhleh L, Ram PT (2008) The signaling petri net-based simulator: a non-parametric strategy for characterizing the dynamics of cell-specific signaling networks. PLoS Comput Biol 4: e1000005.
192. Watterson S, Marshall S, Ghazal P (2008) Logic models of pathway biology. Drug Discov Today 13: 447-456.
193. Raza S, McDerment N, Lacaze PA, Robertson K, Watterson S, et al. (2010) Construction of a large scale integrated map of macrophage pathogen recognition and effector systems. BMC Syst Biol 4: 63.
194. Nurse P (2000) A long twentieth century of the cell cycle and beyond. Cell 100: 71-78.
195. Stillman B (1996) Cell cycle control of DNA replication. Science 274: 1659-1664.
196. Nicklas RB (1997) How cells get the right chromosomes. Science 275: 632-637.
197. Qu Z, MacLellan WR, Weiss JN (2003) Dynamics of the cell cycle: checkpoints, sizers, and timers. Biophys J 85: 3600-3611.
198. Bootsma D, Budke L, Vos O (1964) Studies on Synchronous Division of Tissue Culture Cells Initiated by Excess Thymidine. Exp Cell Res 33: 301-309.
199. Huberman JA (1981) New views of the biochemistry of eucaryotic DNA replication revealed by aphidicolin, an unusual inhibitor of DNA polymerase alpha. Cell 23: 647-648.

200. Zieve GW, Turnbull D, Mullins JM, McIntosh JR (1980) Production of large numbers of mitotic mammalian cells by use of the reversible microtubule inhibitor nocodazole. Nocodazole accumulated mitotic cells. Exp Cell Res 126: 397-405.
201. Schneiderman MH, Dewey WC, Leeper DB, Nagasawa H (1972) Use of the mitotic selection procedure for cell cycle analysis. Comparison between the X-ray and cycloheximide G2 markers. Exp Cell Res 74: 430-438.
202. Ali SH, DeCaprio JA (2001) Cellular transformation by SV40 large T antigen: interaction with host proteins. Semin Cancer Biol 11: 15-23.
203. DuBridge RB, Tang P, Hsia HC, Leong PM, Miller JH, et al. (1987) Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. Mol Cell Biol 7: 379-387.
204. Pavelin J, Reynolds N, Chiweshe S, Wu G, Tiribassi R, et al. (2013) Systematic microRNA analysis identifies ATP6V0C as an essential host factor for human cytomegalovirus replication. PLoS Pathog 9: e1003820.
205. Affymetrix (2015) GeneTitan® Multi-Channel (MC).
206. Affymetrix (2015) GeneAtlas® Personal Microarray System.
207. Carvalho BS, Irizarry RA (2010) A framework for oligonucleotide microarray preprocessing. Bioinformatics 26: 2363-2367.
208. Team RC (2013) R: A Language and Environment for Statistical Computing.
209. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249-264.
210. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D (1997) GeneCards: integrating information about genes, proteins and diseases. Trends Genet 13: 163.
211. (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Res 41: D43-47.
212. Huang da W, Sherman BT, Tan Q, Collins JR, Alvord WG, et al. (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol 8: R183.
213. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25-29.
214. Consortium GO (2015) GO Slim and Subset Guide.
215. Consortium F, the RP, Clst, Forrest AR, Kawaji H, et al. (2014) A promoter-level mammalian expression atlas. Nature 507: 462-470.
216. Flicek P, Aken BL, Ballester B, Beal K, Bragin E, et al. (2010) Ensembl's 10th year. Nucleic Acids Res 38: D557-562.
217. Ouyang B, Lan Z, Meadows J, Pan H, Fukasawa K, et al. (1998) Human Bub1: a putative spindle checkpoint kinase closely linked to cell proliferation. Cell Growth Differ 9: 877-885.
218. Roberti R, Bennati AM, Galli G, Caruso D, Maras B, et al. (2002) Cloning and expression of sterol Delta 14-reductase from bovine liver. Eur J Biochem 269: 283-290.
219. Lamour V, Lecluse Y, Desmaze C, Spector M, Bodescot M, et al. (1995) A human homolog of the S. cerevisiae HIR1 and HIR2 transcriptional repressors cloned from the DiGeorge syndrome critical region. Hum Mol Genet 4: 791-799.
220. Lin JX, Mietz J, Modi WS, John S, Leonard WJ (1996) Cloning of human Stat5B. Reconstitution of interleukin-2-induced Stat5A and Stat5B DNA binding activity in COS-7 cells. J Biol Chem 271: 10738-10744.

221. Rasche N, Dybkov O, Schmitzova J, Akyildiz B, Fabrizio P, et al. (2012) Cwc2 and its human homologue RBM22 promote an active conformation of the spliceosome catalytic centre. EMBO J 31: 1591-1604.
222. Bashir MM, Han MD, Abrams WR, Tucker T, Ma RI, et al. (1996) Analysis of the human gene encoding latent transforming growth factor-beta-binding protein-2. Int J Biochem Cell Biol 28: 531-542.
223. Oda Y, Okada T, Yoshida H, Kaufman RJ, Nagata K, et al. (2006) Derlin-2 and Derlin-3 are regulated by the mammalian unfolded protein response and are required for ER-associated degradation. J Cell Biol 172: 383-393.
224. Kumagai H, Sato N, Yamada M, Mahony D, Seghezzi W, et al. (1999) A novel growth- and cell cycle-regulated protein, ASK, activates human Cdc7-related kinase and is essential for G1/S transition in mammalian cells. Mol Cell Biol 19: 5083-5095.
225. Chini CC, Chen J (2003) Human claspin is required for replication checkpoint control. J Biol Chem 278: 30057-30062.
226. Miller KA, Yoshikawa DM, McConnell IR, Clark R, Schild D, et al. (2002) RAD51C interacts with RAD51B and is central to a larger protein complex in vivo exclusive of RAD51. J Biol Chem 277: 8406-8411.
227. Rubio ED, Reiss DJ, Welcsh PL, Disteche CM, Filippova GN, et al. (2008) CTCF physically links cohesin to chromatin. Proc Natl Acad Sci U S A 105: 8309-8314.
228. Oshimori N, Li X, Ohsugi M, Yamamoto T (2009) Cep72 regulates the localization of key centrosomal proteins and proper bipolar spindle formation. EMBO J 28: 2066-2076.
229. Jiang W, Jimenez G, Wells NJ, Hope TJ, Wahl GM, et al. (1998) PRC1: a human mitotic spindle-associated CDK substrate protein required for cytokinesis. Mol Cell 2: 877-885.
230. Harper JW, Adami GR, Wei N, Keyomarsi K, Elledge SJ (1993) The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. Cell 75: 805-816.
231. Hannon GJ, Beach D (1994) p15INK4B is a potential effector of TGF-beta-induced cell cycle arrest. Nature 371: 257-261.
232. Lew DJ, Dulic V, Reed SI (1991) Isolation of three novel human cyclins by rescue of G1 cyclin (Cln) function in yeast. Cell 66: 1197-1206.
233. Meyerson M, Harlow E (1994) Identification of G1 kinase activity for cdk6, a novel cyclin D partner. Mol Cell Biol 14: 2077-2086.
234. Sardet C, Vidal M, Cobrinik D, Geng Y, Onufryk C, et al. (1995) E2F-4 and E2F-5, two members of the E2F family, are expressed in the early phases of the cell cycle. Proc Natl Acad Sci U S A 92: 2403-2407.
235. Zariwala M, Liu J, Xiong Y (1998) Cyclin E2, a novel human G1 cyclin and activating partner of CDK2 and CDK3, is induced by viral oncoproteins. Oncogene 17: 2787-2798.
236. Blomberg I, Hoffmann I (1999) Ectopic expression of Cdc25A accelerates the G(1)/S transition and leads to premature activation of cyclin E- and cyclin A-dependent kinases. Mol Cell Biol 19: 6183-6194.
237. Reardon JT, Ge H, Gibbs E, Sancar A, Hurwitz J, et al. (1996) Isolation and characterization of two human transcription factor IIH (TFIIH)-related complexes: ERCC2/CAK and TFIIH. Proc Natl Acad Sci U S A 93: 6482-6487.
238. Saha P, Chen J, Thome KC, Lawlis SJ, Hou ZH, et al. (1998) Human CDC6/Cdc18 associates with Orc1 and cyclin-cdk and is selectively eliminated from the nucleus at the onset of S phase. Mol Cell Biol 18: 2758-2767.

239. Ritzi M, Baack M, Musahl C, Romanowski P, Laskey RA, et al. (1998) Human minichromosome maintenance proteins and human origin recognition complex 2 protein on chromatin. J Biol Chem 273: 24543-24549.
240. Guenebeaud C, Goldschneider D, Castets M, Guix C, Chazot G, et al. (2010) The dependence receptor UNC5H2/B triggers apoptosis via PP2A-mediated dephosphorylation of DAP kinase. Mol Cell 40: 863-876.
241. Watt PM, Hickson ID (1994) Structure and function of type II DNA topoisomerases. Biochem J 303 (Pt 3): 681-695.
242. Pagano M, Pepperkok R, Verde F, Ansorge W, Draetta G (1992) Cyclin A is required at two points in the human cell cycle. EMBO J 11: 961-971.
243. Bellanger S, de Gramont A, Sobczak-Thepot J (2007) Cyclin B2 suppresses mitotic failure and DNA re-replication in human somatic cells knocked down for both cyclins B1 and B2. Oncogene 26: 7175-7184.
244. Lindqvist A, Kallstrom H, Lundgren A, Barsoum E, Rosenthal CK (2005) Cdc25B cooperates with Cdc25A to induce mitosis but has a unique role in activating cyclin B1-Cdk1 at the centrosome. J Cell Biol 171: 35-45.
245. Stegmeier F, Rape M, Draviam VM, Nalepa G, Sowa ME, et al. (2007) Anaphase initiation is regulated by antagonistic ubiquitination and deubiquitination activities. Nature 446: 876-881.
246. Nakaya T, Kuwahara K, Ohta K, Kitabatake M, Toda T, et al. (2010) Critical role of Pcid2 in B cell survival through the regulation of MAD2 expression. J Immunol 185: 5180-5187.
247. Bomont P, Maddox P, Shah JV, Desai AB, Cleveland DW (2005) Unstable microtubule capture at kinetochores depleted of the centromere-associated protein CENP-F. EMBO J 24: 3927-3939.
248. Zou JX, Revenko AS, Li LB, Gemo AT, Chen HW (2007) ANCCA, an estrogen-regulated AAA+ ATPase coactivator for ERalpha, is required for coregulator occupancy and chromatin modification. Proc Natl Acad Sci U S A 104: 18067-18072.
249. Tullio-Pelet A, Salomon R, Hadj-Rabia S, Mugnier C, de Laet MH, et al. (2000) Mutant WD-repeat protein in triple-A syndrome. Nat Genet 26: 332-335.
250. Hanisch A, Sillje HH, Nigg EA (2006) Timely anaphase onset requires a novel spindle and kinetochore complex comprising Ska1 and Ska2. EMBO J 25: 5504-5515.
251. Tanenbaum ME, Macurek L, Janssen A, Geers EF, Alvarez-Fernandez M, et al. (2009) Kif15 cooperates with eg5 to promote bipolar spindle assembly. Curr Biol 19: 1703-1711.
252. Hendzel MJ, Lever MA, Crawford E, Th'ng JP (2004) The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo. J Biol Chem 279: 20028-20034.
253. Collins NB, Wilson JB, Bush T, Thomashevski A, Roberts KJ, et al. (2009) ATR-dependent phosphorylation of FANCA on serine 1449 after DNA damage is important for FA pathway function. Blood 113: 2181-2190.
254. Thompson LH, Brookman KW, Jones NJ, Allen SA, Carrano AV (1990) Molecular cloning of the human XRCC1 gene, which corrects defective DNA strand break repair and sister chromatid exchange. Mol Cell Biol 10: 6160-6171.
255. Harvey CB, Wang Y, Darmoul D, Phillips A, Mantei N, et al. (1996) Characterisation of a human homologue of a yeast cell division cycle gene, MCM6, located adjacent to the 5' end of the lactase gene on chromosome 2q21. FEBS Lett 398: 135-140.
256. Wong SW, Wahl AF, Yuan PM, Arai N, Pearson BE, et al. (1988) Human DNA polymerase alpha gene expression is cell proliferation dependent and its primary

structure is similar to both prokaryotic and eukaryotic replicative DNA polymerases. EMBO J 7: 37-47.
257. Lawo S, Bashkurov M, Mullin M, Ferreria MG, Kittler R, et al. (2009) HAUS, the 8-subunit human Augmin complex, regulates centrosome and spindle integrity. Curr Biol 19: 816-826.
258. Lin YC, Chang CW, Hsu WB, Tang CJ, Lin YN, et al. (2013) Human microcephaly protein CEP135 binds to hSAS-6 and CPAP, and is required for centriole assembly. EMBO J 32: 1141-1154.
259. Muhlenbein N, Hofmann S, Rothbauer U, Bauer MF (2004) Organization and function of the small Tim complexes acting along the import pathway of metabolite carriers into mammalian mitochondria. J Biol Chem 279: 13540-13546.
260. Campisi J, Morreo G, Pardee AB (1984) Kinetics of G1 transit following brief starvation for serum factors. Exp Cell Res 152: 459-466.
261. Whitfield ML, George LK, Grant GD, Perou CM (2006) Common markers of proliferation. Nat Rev Cancer 6: 99-106.
262. Gong J, Traganos F, Darzynkiewicz Z (1995) Growth imbalance and altered expression of cyclins B1, A, E, and D3 in MOLT-4 cells synchronized in the cell cycle by inhibitors of DNA replication. Cell Growth Differ 6: 1485-1493.
263. Kung AL, Zetterberg A, Sherwood SW, Schimke RT (1990) Cytotoxic effects of cell cycle phase specific agents: result of cell cycle perturbation. Cancer Res 50: 7307-7317.
264. Doyon Y, Cayrou C, Ullah M, Landry AJ, Cote V, et al. (2006) ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. Mol Cell 21: 51-64.
265. Hu W, Feng Z, Levine AJ (2012) The Regulation of Multiple p53 Stress Responses is Mediated through MDM2. Genes Cancer 3: 199-208.
266. Mandl A, Sarkes D, Carricaburu V, Jung V, Rameh L (2007) Serum withdrawal-induced accumulation of phosphoinositide 3-kinase lipids in differentiating 3T3-L6 myoblasts: distinct roles for Ship2 and PTEN. Mol Cell Biol 27: 8098-8112.
267. Tobey RA, Valdez JG, Crissman HA (1988) Synchronization of human diploid fibroblasts at multiple stages of the cell cycle. Exp Cell Res 179: 400-416.
268. Pardee AB (1989) G1 events and regulation of cell proliferation. Science 246: 603-608.
269. Winkles JA (1998) Serum- and polypeptide growth factor-inducible gene expression in mouse fibroblasts. Prog Nucleic Acid Res Mol Biol 58: 41-78.
270. Lukinavicius G, Lavogina D, Orpinell M, Umezawa K, Reymond L, et al. (2013) Selective chemical crosslinking reveals a Cep57-Cep63-Cep152 centrosomal complex. Curr Biol 23: 265-270.
271. Feng L, Huang J, Chen J (2009) MERIT40 facilitates BRCA1 localization and DNA damage repair. Genes Dev 23: 719-728.
272. Zhuo D, Zhao WD, Wright FA, Yang HY, Wang JP, et al. (2001) Assembly, annotation, and integration of UNIGENE clusters into the human genome draft. Genome Res 11: 904-918.
273. (2011) Ongoing and future developments at the Universal Protein Resource. Nucleic Acids Res 39: D214-219.
274. Zhang JH, Chung TDY, Oldenburg KR (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. Journal of Biomolecular Screening 4: 67-73.
275. Asphahani F, Zhang M (2007) Cellular impedance biosensors for drug screening and toxin detection. Analyst 132: 835-841.
276. Chesnoy S, Huang L (2000) Structure and function of lipid-DNA complexes for gene delivery. Annu Rev Biophys Biomol Struct 29: 27-47.

277. Nicolau C, Legrand A, Grosse E (1987) Liposomes as carriers for in vivo gene transfer and expression. Methods Enzymol 149: 157-176.
278. Plank C, Schillinger U, Scherer F, Bergemann C, Remy JS, et al. (2003) The magnetofection method: using magnetic force to enhance gene delivery. Biol Chem 384: 737-747.
279. Kittler R, Surendranath V, Heninger AK, Slabicki M, Theis M, et al. (2007) Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. Nat Methods 4: 337-344.
280. Kittler R, Heninger AK, Franke K, Habermann B, Buchholz F (2005) Production of endoribonuclease-prepared short interfering RNAs for gene silencing in mammalian cells. Nat Methods 2: 779-784.
281. Chen Q, Zhang X, Jiang Q, Clarke PR, Zhang C (2008) Cyclin B1 is localized to unattached kinetochores and contributes to efficient microtubule attachment and proper chromosome alignment during mitosis. Cell Res 18: 268-280.
282. Rossert J, Terraz C, Dupont S (2000) Regulation of type I collagen genes expression. Nephrol Dial Transplant 15 Suppl 6: 66-68.
283. Zhang JD (2012) RTCA: Open-source toolkit to analyse data from xCELLigence System (RTCA).
284. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5.
285. R Development Core Team (2008) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
286. Wilk MB, Gnanadesikan R (1968) Probability plotting methods for the analysis of data. Biometrika 55: 1-17.
287. Boutros M, Bras LP, Huber W (2006) Analysis of cell-based RNAi screens. Genome Biol 7.
288. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, et al. (2010) Towards a knowledge-based Human Protein Atlas. Nat Biotechnol 28: 1248-1250.
289. Campbell PI (1983) Toxicity of some charged lipids used in liposome preparations. Cytobios 37: 21-26.
290. Mayhew E, Ito M, Lazo R (1987) Toxicity of non-drug-containing liposomes for cultured human cells. Exp Cell Res 171: 195-202.
291. Romoren K, Thu BJ, Bols NC, Evensen O (2004) Transfection efficiency and cytotoxicity of cationic liposomes in salmonid cell lines of hepatocyte and macrophage origin. Biochim Biophys Acta 1663: 127-134.
292. Yen HC, Xu Q, Chou DM, Zhao Z, Elledge SJ (2008) Global protein stability profiling in mammalian cells. Science 322: 918-923.
293. Mendez J (2003) Cell proliferation without cyclin E-CDK2. Cell 114: 398-399.
294. Geng Y, Yu Q, Sicinska E, Das M, Schneider JE, et al. (2003) Cyclin E ablation in the mouse. Cell 114: 431-443.
295. Boutros M, Ahringer J (2008) The art and design of genetic screens: RNA interference. Nat Rev Genet 9: 554-566.
296. Schlaitz AL, Thompson J, Wong CC, Yates JR, 3rd, Heald R (2013) REEP3/4 ensure endoplasmic reticulum clearance from metaphase chromatin and proper nuclear envelope architecture. Dev Cell 26: 315-323.
297. Obeidat M, Miller S, Probert K, Billington CK, Henry AP, et al. (2013) GSTCD and INTS12 regulation and expression in the human lung. PLoS One 8: e74630.
298. Zimmerman SW, Yi YJ, Sutovsky M, van Leeuwen FW, Conant G, et al. (2014) Identification and characterization of RING-finger ubiquitin ligase UBR7 in mammalian spermatozoa. Cell Tissue Res 356: 261-278.

299. Jain M, Zhang L, Boufraqech M, Liu-Chittenden Y, Bussey K, et al. (2014) ZNF367 inhibits cancer progression and is targeted by miR-195. PLoS One 9: e101423.
300. De Souza Santos E, De Bessa SA, Netto MM, Nagai MA (2008) Silencing of LRRC49 and THAP10 genes by bidirectional promoter hypermethylation is a frequent event in breast cancer. Int J Oncol 33: 25-31.
301. Calera MR, Zamora-Ramos C, Araiza-Villanueva MG, Moreno-Aguilar CA, Pena-Gomez SG, et al. (2011) Parcs/Gpn3 is required for the nuclear accumulation of RNA polymerase II. Biochim Biophys Acta 1813: 1708-1716.
302. Isojima T, Doi K, Mitsui J, Oda Y, Tokuhiro E, et al. (2014) A recurrent de novo FAM111A mutation causes Kenny-Caffey syndrome type 2. J Bone Miner Res 29: 992-998.
303. Xu J, Zhou X, Wang J, Li Z, Kong X, et al. (2013) RhoGAPs attenuate cell proliferation by direct interaction with p53 tetramerization domain. Cell Rep 3: 1526-1538.
304. Rismanchi N, Soderblom C, Stadler J, Zhu PP, Blackstone C (2008) Atlastin GTPases are required for Golgi apparatus and ER morphogenesis. Hum Mol Genet 17: 1591-1604.
305. Wong QW, Li J, Ng SR, Lim SG, Yang H, et al. (2014) RPL39L is an example of a recently evolved ribosomal protein paralog that shows highly specific tissue expression patterns and is upregulated in ESCs and HCC tumors. RNA Biol 11: 33-41.
306. Florio M, Albert M, Taverna E, Namba T, Brandl H, et al. (2015) Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. Science 347: 1465-1470.
307. Marchesi S, Montani F, Deflorian G, D'Antuono R, Cuomo A, et al. (2014) DEPDC1B coordinates de-adhesion events and cell-cycle progression at mitosis. Dev Cell 31: 420-433.
308. Yuan J, Chen J (2013) FIGNL1-containing protein complex is required for efficient homologous recombination repair. Proc Natl Acad Sci U S A 110: 10640-10645.
309. Aydin H, Azimi FC, Cook JD, Lee JE (2012) A convenient and general expression platform for the production of secreted proteins from human cells. J Vis Exp.
310. Thomas P, Smart TG (2005) HEK293 cell line: a vehicle for the expression of recombinant proteins. J Pharmacol Toxicol Methods 51: 187-200.
311. Yavin E, Yavin Z (1974) Attachment and culture of dissociated cells from rat embryo cerebral hemispheres on polylysine-coated surface. J Cell Biol 62: 540-546.
312. Oegema K, Savoian MS, Mitchison TJ, Field CM (2000) Functional analysis of a human homologue of the Drosophila actin binding protein anillin suggests a role in cytokinesis. J Cell Biol 150: 539-552.
313. Mi Y, Zhang C, Bu Y, Zhang Y, He L, et al. (2015) DEPDC1 is a novel cell cycle related gene that regulates mitotic progression. BMB Rep 48: 413-418.
314. Simpson F, Lammerts van Bueren K, Butterfield N, Bennetts JS, Bowles J, et al. (2006) The PCNA-associated factor KIAA0101/p15(PAF) binds the potential tumor suppressor product p33ING1b. Exp Cell Res 312: 73-85.
315. Fang L, Seki A, Fang G (2009) SKAP associates with kinetochores and promotes the metaphase-to-anaphase transition. Cell Cycle 8: 2819-2827.
316. Krautkramer KA, Linnemann AK, Fontaine DA, Whillock AL, Harris TW, et al. (2013) Tcf19 is a novel islet factor necessary for proliferation and survival in the INS-1 beta-cell line. Am J Physiol Endocrinol Metab 305: E600-610.
317. Stadler C, Rexhepaj E, Singan VR, Murphy RF, Pepperkok R, et al. (2013) Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. Nat Methods 10: 315-323.

318. Uzbekov RE (2004) Analysis of the cell cycle and a method employing synchronized cells for study of protein expression at various stages of the cell cycle. Biochemistry (Mosc) 69: 485-496.
319. Park SJ, Kim SJ, Rhee Y, Byun JH, Kim SH, et al. (2007) Fidgetin-like 1 gene inhibited by basic fibroblast growth factor regulates the proliferation and differentiation of osteoblasts. J Bone Miner Res 22: 889-896.
320. Fine DA, Rozenblatt-Rosen O, Padi M, Korkhin A, James RL, et al. (2012) Identification of FAM111A as an SV40 host range restriction and adenovirus helper factor. PLoS Pathog 8: e1002949.
321. Tao R, Xu X, Sun C, Wang Y, Wang S, et al. (2015) KPNA2 interacts with P65 to modulate catabolic events in osteoarthritis. Exp Mol Pathol 99: 245-252.
322. Huang L, Wang HY, Li JD, Wang JH, Zhou Y, et al. (2013) KPNA2 promotes cell proliferation and tumorigenicity in epithelial ovarian carcinoma through upregulation of c-Myc and downregulation of FOXO3a. Cell Death Dis 4: e745.
323. Nadano D, Notsu T, Matsuda T, Sato T (2002) A human gene encoding a protein homologous to ribosomal protein L39 is normally expressed in the testis and derepressed in multiple cancer cells. Biochim Biophys Acta 1577: 430-436.
324. Gonczy P (2012) Towards a molecular architecture of centriole assembly. Nat Rev Mol Cell Biol 13: 425-435.
325. Stearns T (2001) Centrosome duplication. a centriolar pas de deux. Cell 105: 417-420.
326. Schmidt KN, Kuhns S, Neuner A, Hub B, Zentgraf H, et al. (2012) Cep164 mediates vesicular docking to the mother centriole during early steps of ciliogenesis. J Cell Biol 199: 1083-1101.
327. Graser S, Stierhof YD, Lavoie SB, Gassner OS, Lamla S, et al. (2007) Cep164, a novel centriole appendage protein required for primary cilium formation. J Cell Biol 179: 321-330.
328. Bettencourt-Dias M, Glover DM (2007) Centrosome biogenesis and function: centrosomics brings new understanding. Nat Rev Mol Cell Biol 8: 451-463.
329. Bhalla US, Iyengar R (1999) Emergent properties of networks of biological signaling pathways. Science 283: 381-387.
330. Klukas C, Schreiber F (2007) Dynamic exploration and editing of KEGG pathway diagrams. Bioinformatics 23: 344-350.
331. Kutmon M, Lotia S, Evelo CT, Pico AR (2014) WikiPathways App for Cytoscape: Making biological pathways amenable to network analysis and visualization. F1000Res 3: 152.
332. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, et al. (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41: D808-815.
333. Gray KA, Yates B, Seal RL, Wright MW, Bruford EA (2014) Genenames.org: the HGNC resources in 2015. Nucleic Acids Res.
334. Bornens M (2002) Centrosome composition and microtubule anchoring mechanisms. Curr Opin Cell Biol 14: 25-34.
335. Oakley BR (2000) An abundance of tubulins. Trends Cell Biol 10: 537-542.
336. Sillibourne JE, Bornens M (2010) Polo-like kinase 4: the odd one out of the family. Cell Div 5: 25.
337. Nogales-Cadenas R, Abascal F, Diez-Perez J, Carazo JM, Pascual-Montano A (2009) CentrosomeDB: a human centrosomal proteins database. Nucleic Acids Res 37: D175-180.
338. Ren J, Liu Z, Gao X, Jin C, Ye M, et al. (2010) MiCroKit 3.0: an integrated database of midbody, centrosome and kinetochore. Nucleic Acids Res 38: D155-160.

339. Hatzopoulos GN, Erat MC, Cutts E, Rogala KB, Slater LM, et al. (2013) Structural analysis of the G-box domain of the microcephaly protein CPAP suggests a role in centriole architecture. Structure 21: 2069-2077.
340. Strnad P, Leidel S, Vinogradova T, Euteneuer U, Khodjakov A, et al. (2007) Regulated HsSAS-6 levels ensure formation of a single procentriole per centriole during the centrosome duplication cycle. Dev Cell 13: 203-213.
341. Guichard P, Hachet V, Majubu N, Neves A, Demurtas D, et al. (2013) Native architecture of the centriole proximal region reveals features underlying its 9-fold radial symmetry. Curr Biol 23: 1620-1628.
342. Paintrand M, Moudjou M, Delacroix H, Bornens M (1992) Centrosome organization and centriole architecture: their sensitivity to divalent cations. J Struct Biol 108: 107-128.
343. Hung LY, Tang CJ, Tang TK (2000) Protein 4.1 R-135 interacts with a novel centrosomal protein (CPAP) which is associated with the gamma-tubulin complex. Mol Cell Biol 20: 7813-7825.
344. Chen Z, Indjeian VB, McManus M, Wang L, Dynlacht BD (2002) CP110, a cell cycle-dependent CDK substrate, regulates centrosome duplication in human cells. Dev Cell 3: 339-350.
345. Li J, D'Angiolella V, Seeley ES, Kim S, Kobayashi T, et al. (2013) USP33 regulates centrosome biogenesis via deubiquitination of the centriolar protein CP110. Nature 495: 255-259.
346. Venoux M, Tait X, Hames RS, Straatman KR, Woodland HR, et al. (2013) Poc1A and Poc1B act together in human cells to ensure centriole integrity. J Cell Sci 126: 163-175.
347. Azimzadeh J, Hergert P, Delouvee A, Euteneuer U, Formstecher E, et al. (2009) hPOC5 is a centrin-binding protein required for assembly of full-length centrioles. J Cell Biol 185: 101-114.
348. Singla V, Romaguera-Ros M, Garcia-Verdugo JM, Reiter JF (2010) Ofd1, a human disease gene, regulates the length and distal structure of centrioles. Dev Cell 18: 410-424.
349. Soung NK, Kang YH, Kim K, Kamijo K, Yoon H, et al. (2006) Requirement of hCenexin for proper mitotic functions of polo-like kinase 1 at the centrosomes. Mol Cell Biol 26: 8316-8335.
350. Levy YY, Lai EY, Remillard SP, Heintzelman MB, Fulton C (1996) Centrin is a conserved protein that forms diverse associations with centrioles and MTOCs in Naegleria and other organisms. Cell Motil Cytoskeleton 33: 298-323.
351. Salisbury JL (1995) Centrin, centrosomes, and mitotic spindle poles. Curr Opin Cell Biol 7: 39-45.
352. Lutz W, Lingle WL, McCormick D, Greenwood TM, Salisbury JL (2001) Phosphorylation of centrin during the cell cycle and its role in centriole separation preceding centrosome duplication. J Biol Chem 276: 20774-20780.
353. Mennella V, Keszthelyi B, McDonald KL, Chhun B, Kan F, et al. (2012) Subdiffraction-resolution fluorescence microscopy reveals a domain of the centrosome critical for pericentriolar material organization. Nat Cell Biol 14: 1159-1168.
354. Gunawardane RN, Lizarraga SB, Wiese C, Wilde A, Zheng Y (2000) gamma-Tubulin complexes and their role in microtubule nucleation. Curr Top Dev Biol 49: 55-73.
355. Gunawardane RN, Martin OC, Cao K, Zhang L, Dej K, et al. (2000) Characterization and reconstitution of Drosophila gamma-tubulin ring complex subunits. J Cell Biol 151: 1513-1524.

356. Murphy SM, Preble AM, Patel UK, O'Connell KL, Dias DP, et al. (2001) GCP5 and GCP6: two new members of the human gamma-tubulin complex. Mol Biol Cell 12: 3340-3352.
357. Gunawardane RN, Martin OC, Zheng Y (2003) Characterization of a new gammaTuRC subunit with WD repeats. Mol Biol Cell 14: 1017-1026.
358. Haren L, Remy MH, Bazin I, Callebaut I, Wright M, et al. (2006) NEDD1-dependent recruitment of the gamma-tubulin ring complex to the centrosome is necessary for centriole duplication and spindle assembly. J Cell Biol 172: 505-515.
359. Luders J, Patel UK, Stearns T (2006) GCP-WD is a gamma-tubulin targeting factor required for centrosomal and chromatin-mediated microtubule nucleation. Nat Cell Biol 8: 137-147.
360. Teixido-Travesa N, Villen J, Lacasa C, Bertran MT, Archinti M, et al. (2010) The gammaTuRC revisited: a comparative analysis of interphase and mitotic human gammaTuRC redefines the set of core components and identifies the novel subunit GCP8. Mol Biol Cell 21: 3963-3972.
361. Manning J, Kumar S (2007) NEDD1: function in microtubule nucleation, spindle assembly and beyond. Int J Biochem Cell Biol 39: 7-11.
362. Manning JA, Shalini S, Risk JM, Day CL, Kumar S (2010) A direct interaction with NEDD1 regulates gamma-tubulin recruitment to the centrosome. PLoS One 5: e9618.
363. Zeng CJ, Lee YR, Liu B (2009) The WD40 repeat protein NEDD1 functions in microtubule organization during cell division in Arabidopsis thaliana. Plant Cell 21: 1129-1140.
364. Dhani DK, Goult BT, George GM, Rogerson DT, Bitton DA, et al. (2013) Mzt1/Tam4, a fission yeast MOZART1 homologue, is an essential component of the gamma-tubulin complex and directly interacts with GCP3(Alp6). Mol Biol Cell 24: 3337-3349.
365. Dammermann A, Merdes A (2002) Assembly of centrosomal proteins and microtubule organization depends on PCM-1. J Cell Biol 159: 255-266.
366. Takahashi M, Yamagiwa A, Nishimura T, Mukai H, Ono Y (2002) Centrosomal proteins CG-NAP and kendrin provide microtubule nucleation sites by anchoring gamma-tubulin ring complex. Mol Biol Cell 13: 3235-3245.
367. Zimmerman WC, Sillibourne J, Rosa J, Doxsey SJ (2004) Mitosis-specific anchoring of gamma tubulin complexes by pericentrin controls spindle organization and mitotic entry. Mol Biol Cell 15: 3642-3657.
368. Flory MR, Moser MJ, Monnat RJ, Jr., Davis TN (2000) Identification of a human centrosomal calmodulin-binding protein that shares homology with pericentrin. Proc Natl Acad Sci U S A 97: 5919-5923.
369. Gillingham AK, Munro S (2000) The PACT domain, a conserved centrosomal targeting motif in the coiled-coil proteins AKAP450 and pericentrin. EMBO Rep 1: 524-529.
370. Fong KW, Choi YK, Rattner JB, Qi RZ (2008) CDK5RAP2 is a pericentriolar protein that functions in centrosomal attachment of the gamma-tubulin ring complex. Mol Biol Cell 19: 115-125.
371. Gopalakrishnan J, Mennella V, Blachon S, Zhai B, Smith AH, et al. (2011) Sas-4 provides a scaffold for cytoplasmic complexes and tethers them in a centrosome. Nat Commun 2: 359.
372. Brown NJ, Marjanovic M, Luders J, Stracker TH, Costanzo V (2013) Cep63 and cep152 cooperate to ensure centriole duplication. PLoS One 8: e69986.
373. Blachon S, Gopalakrishnan J, Omori Y, Polyanovsky A, Church A, et al. (2008) Drosophila asterless and vertebrate Cep152 Are orthologs essential for centriole duplication. Genetics 180: 2081-2094.

374. Lin YN, Wu CT, Lin YC, Hsu WB, Tang CJ, et al. (2013) CEP120 interacts with CPAP and positively regulates centriole elongation. J Cell Biol 202: 211-219.
375. Winey M, O'Toole E (2014) Centriole structure. Philos Trans R Soc Lond B Biol Sci 369.
376. Jana SC, Marteil G, Bettencourt-Dias M (2014) Mapping molecules to structure: unveiling secrets of centriole and cilia assembly with near-atomic resolution. Curr Opin Cell Biol 26: 96-106.
377. Mogensen MM, Malik A, Piel M, Bouckson-Castaing V, Bornens M (2000) Microtubule minus-end anchorage at centrosomal and non-centrosomal sites: the role of ninein. J Cell Sci 113 (Pt 17): 3013-3023.
378. Bouckson-Castaing V, Moudjou M, Ferguson DJ, Mucklow S, Belkaid Y, et al. (1996) Molecular characterisation of ninein, a new coiled-coil protein of the centrosome. J Cell Sci 109 (Pt 1): 179-190.
379. Ou YY, Mack GJ, Zhang M, Rattner JB (2002) CEP110 and ninein are located in a specific domain of the centrosome associated with centrosome maturation. J Cell Sci 115: 1825-1835.
380. Chang P, Giddings TH, Jr., Winey M, Stearns T (2003) Epsilon-tubulin is required for centriole duplication and microtubule organization. Nat Cell Biol 5: 71-76.
381. Kobayashi T, Tsang WY, Li J, Lane W, Dynlacht BD (2011) Centriolar kinesin Kif24 interacts with CP110 to remodel microtubules and regulate ciliogenesis. Cell 145: 914-925.
382. Su LK, Qi Y (2001) Characterization of human MAPRE genes and their proteins. Genomics 71: 142-149.
383. Guarguaglini G, Duncan PI, Stierhof YD, Holmstrom T, Duensing S, et al. (2005) The forkhead-associated domain protein Cep170 interacts with Polo-like kinase 1 and serves as a marker for mature centrioles. Mol Biol Cell 16: 1095-1107.
384. Welburn JP, Cheeseman IM (2012) The microtubule-binding protein Cep170 promotes the targeting of the kinesin-13 depolymerase Kif2b to the mitotic spindle. Mol Biol Cell 23: 4786-4795.
385. Cizmecioglu O, Warnke S, Arnold M, Duensing S, Hoffmann I (2008) Plk2 regulated centriole duplication is dependent on its localization to the centrioles and a functional polo-box domain. Cell Cycle 7: 3548-3555.
386. Warnke S, Kemmler S, Hames RS, Tsai HL, Hoffmann-Rohrer U, et al. (2004) Polo-like kinase-2 is required for centriole duplication in mammalian cells. Curr Biol 14: 1200-1207.
387. Sluder G, Khodjakov A (2010) Centriole duplication: analogue control in a digital age. Cell Biol Int 34: 1239-1245.
388. Strnad P, Gonczy P (2008) Mechanisms of procentriole formation. Trends Cell Biol 18: 389-396.
389. Kleylein-Sohn J, Westendorf J, Le Clech M, Habedanck R, Stierhof YD, et al. (2007) Plk4-induced centriole biogenesis in human cells. Dev Cell 13: 190-202.
390. Kirkham M, Muller-Reichert T, Oegema K, Grill S, Hyman AA (2003) SAS-4 is a C. elegans centriolar protein that controls centrosome size. Cell 112: 575-587.
391. Leidel S, Gonczy P (2003) SAS-4 is essential for centrosome duplication in C elegans and is recruited to daughter centrioles once per cell cycle. Dev Cell 4: 431-439.
392. Kohlmaier G, Loncarek J, Meng X, McEwen BF, Mogensen MM, et al. (2009) Overly long centrioles and defective cell division upon excess of the SAS-4-related protein CPAP. Curr Biol 19: 1012-1018.
393. Pelletier L, O'Toole E, Schwager A, Hyman AA, Muller-Reichert T (2006) Centriole assembly in Caenorhabditis elegans. Nature 444: 619-623.

394. Cizmecioglu O, Arnold M, Bahtz R, Settele F, Ehret L, et al. (2010) Cep152 acts as a scaffold for recruitment of Plk4 and CPAP to the centrosome. J Cell Biol 191: 731-739.
395. Dzhindzhev NS, Yu QD, Weiskopf K, Tzolovsky G, Cunha-Ferreira I, et al. (2010) Asterless is a scaffold for the onset of centriole assembly. Nature 467: 714-718.
396. Hatch EM, Kulukian A, Holland AJ, Cleveland DW, Stearns T (2010) Cep152 interacts with Plk4 and is required for centriole duplication. J Cell Biol 191: 721-729.
397. Ohta M, Ashikawa T, Nozaki Y, Kozuka-Hata H, Goto H, et al. (2014) Direct interaction of Plk4 with STIL ensures formation of a single procentriole per parental centriole. Nat Commun 5: 5267.
398. Sonnen KF, Gabryjonczyk AM, Anselm E, Stierhof YD, Nigg EA (2013) Human Cep192 and Cep152 cooperate in Plk4 recruitment and centriole duplication. J Cell Sci 126: 3223-3233.
399. Puklowski A, Homsi Y, Keller D, May M, Chauhan S, et al. (2011) The SCF-FBXW5 E3-ubiquitin ligase is regulated by PLK4 and targets HsSAS-6 to control centrosome duplication. Nat Cell Biol 13: 1004-1009.
400. Delattre M, Leidel S, Wani K, Baumer K, Bamat J, et al. (2004) Centriolar SAS-5 is required for centrosome duplication in C. elegans. Nat Cell Biol 6: 656-664.
401. Hiraki M, Nakazawa Y, Kamiya R, Hirono M (2007) Bld10p constitutes the cartwheel-spoke tip and stabilizes the 9-fold symmetry of the centriole. Curr Biol 17: 1778-1783.
402. Stevens NR, Dobbelaere J, Brunk K, Franz A, Raff JW (2010) Drosophila Ana2 is a conserved centriole duplication factor. J Cell Biol 188: 313-323.
403. Stevens NR, Roque H, Raff JW (2010) DSas-6 and Ana2 coassemble into tubules to promote centriole duplication and engagement. Dev Cell 19: 913-919.
404. Arquint C, Sonnen KF, Stierhof YD, Nigg EA (2012) Cell-cycle-regulated expression of STIL controls centriole number in human cells. J Cell Sci 125: 1342-1352.
405. Tang CJ, Lin SY, Hsu WB, Lin YN, Wu CT, et al. (2011) The human microcephaly protein STIL interacts with CPAP and is required for procentriole formation. EMBO J 30: 4790-4804.
406. Vulprecht J, David A, Tibelius A, Castiel A, Konotop G, et al. (2012) STIL is required for centriole duplication in human cells. J Cell Sci 125: 1353-1362.
407. Guichard P, Chretien D, Marco S, Tassin AM (2010) Procentriole assembly revealed by cryo-electron tomography. EMBO J 29: 1565-1572.
408. Le Clech M (2008) Role of CAP350 in centriolar tubule stability and centriole assembly. PLoS One 3: e3855.
409. Kollman JM, Polka JK, Zelter A, Davis TN, Agard DA (2010) Microtubule nucleating gamma-TuSC assembles structures with 13-fold microtubule-like symmetry. Nature 466: 879-882.
410. Oegema K, Wiese C, Martin OC, Milligan RA, Iwamatsu A, et al. (1999) Characterization of two related Drosophila gamma-tubulin complexes that differ in their ability to nucleate microtubules. J Cell Biol 144: 721-733.
411. Moritz M, Braunfeld MB, Sedat JW, Alberts B, Agard DA (1995) Microtubule nucleation by gamma-tubulin-containing rings in the centrosome. Nature 378: 638-640.
412. Moritz M, Zheng Y, Alberts BM, Oegema K (1998) Recruitment of the gamma-tubulin ring complex to Drosophila salt-stripped centrosome scaffolds. J Cell Biol 142: 775-786.
413. Murphy SM, Urbani L, Stearns T (1998) The mammalian gamma-tubulin complex contains homologues of the yeast spindle pole body components spc97p and spc98p. J Cell Biol 141: 663-674.

414. Dupuis-Williams P, Fleury-Aubusson A, de Loubresse NG, Geoffroy H, Vayssie L, et al. (2002) Functional role of epsilon-tubulin in the assembly of the centriolar microtubule scaffold. J Cell Biol 158: 1183-1193.
415. Dutcher SK, Morrissette NS, Preble AM, Rackley C, Stanga J (2002) Epsilon-tubulin is an essential component of the centriole. Mol Biol Cell 13: 3859-3869.
416. Agromayor M, Martin-Serrano J (2013) Knowing when to cut and run: mechanisms that control cytokinetic abscission. Trends Cell Biol 23: 433-441.
417. Tsou MF, Wang WJ, George KA, Uryu K, Stearns T, et al. (2009) Polo kinase and separase regulate the mitotic licensing of centriole duplication in human cells. Dev Cell 17: 344-354.
418. Loncarek J, Hergert P, Khodjakov A (2010) Centriole reduplication during prolonged interphase requires procentriole maturation governed by Plk1. Curr Biol 20: 1277-1282.
419. Wang WJ, Soni RK, Uryu K, Tsou MF (2011) The conversion of centrioles to centrosomes: essential coupling of duplication with segregation. J Cell Biol 193: 727-739.
420. Comartin D, Gupta GD, Fussner E, Coyaud E, Hasegan M, et al. (2013) CEP120 and SPICE1 cooperate with CPAP in centriole elongation. Curr Biol 23: 1360-1366.
421. Nakamura A, Arai H, Fujita N (2009) Centrosomal Aki1 and cohesin function in separase-regulated centriole disengagement. J Cell Biol 187: 607-614.
422. Wang G, Jiang Q, Zhang C (2014) The role of mitotic kinases in coupling the centrosome cycle with the assembly of the mitotic spindle. J Cell Sci 127: 4111-4122.
423. Matsuo K, Ohsumi K, Iwabuchi M, Kawamata T, Ono Y, et al. (2012) Kendrin is a novel substrate for separase involved in the licensing of centriole duplication. Curr Biol 22: 915-921.
424. Thein KH, Kleylein-Sohn J, Nigg EA, Gruneberg U (2007) Astrin is required for the maintenance of sister chromatid cohesion and centrosome integrity. J Cell Biol 178: 345-354.
425. Wang X, Yang Y, Duan Q, Jiang N, Huang Y, et al. (2008) sSgo1, a major splice variant of Sgo1, functions in centriole cohesion where it is regulated by Plk1. Dev Cell 14: 331-341.
426. Tsang WY, Dynlacht BD (2008) sSgo1, a guardian of centriole cohesion. Dev Cell 14: 320-322.
427. Tang Z, Shu H, Qi W, Mahmood NA, Mumby MC, et al. (2006) PP2A is required for centromeric localization of Sgo1 and proper chromosome segregation. Dev Cell 10: 575-585.
428. Lopes CA, Prosser SL, Romio L, Hirst RA, O'Callaghan C, et al. (2011) Centriolar satellites are assembly points for proteins implicated in human ciliopathies, including oral-facial-digital syndrome 1. J Cell Sci 124: 600-612.
429. Kim J, Krishnaswami SR, Gleeson JG (2008) CEP290 interacts with the centriolar satellite component PCM-1 and is required for Rab8 localization to the primary cilium. Hum Mol Genet 17: 3796-3805.
430. Kim JC, Badano JL, Sibold S, Esmail MA, Hill J, et al. (2004) The Bardet-Biedl protein BBS4 targets cargo to the pericentriolar region and is required for microtubule anchoring and cell cycle progression. Nat Genet 36: 462-470.
431. Kamiya A, Tan PL, Kubo K, Engelhard C, Ishizuka K, et al. (2008) Recruitment of PCM1 to the centrosome by the cooperative action of DISC1 and BBS4: a candidate for psychiatric illnesses. Arch Gen Psychiatry 65: 996-1006.

432. Kodani A, Tonthat V, Wu B, Sutterlin C (2010) Par6 alpha interacts with the dynactin subunit p150 Glued and is a critical regulator of centrosomal protein recruitment. Mol Biol Cell 21: 3376-3385.
433. Karki S, Holzbaur EL (1999) Cytoplasmic dynein and dynactin in cell division and intracellular transport. Curr Opin Cell Biol 11: 45-53.
434. Ge X, Frank CL, Calderon de Anda F, Tsai LH (2010) Hook3 interacts with PCM1 to regulate pericentriolar material assembly and the timing of neurogenesis. Neuron 65: 191-203.
435. Sedjai F, Acquaviva C, Chevrier V, Chauvin JP, Coppin E, et al. (2010) Control of ciliogenesis by FOR20, a novel centrosome and pericentriolar satellite protein. J Cell Sci 123: 2391-2401.
436. Okuda M (2002) The role of nucleophosmin in centrosome duplication. Oncogene 21: 6170-6174.
437. Okuda M, Horn HF, Tarapore P, Tokuyama Y, Smulian AG, et al. (2000) Nucleophosmin/B23 is a target of CDK2/cyclin E in centrosome duplication. Cell 103: 127-140.
438. Bornens M, Paintrand M, Berges J, Marty MC, Karsenti E (1987) Structural and chemical characterization of isolated centrosomes. Cell Motil Cytoskeleton 8: 238-249.
439. Fry AM, Arnaud L, Nigg EA (1999) Activity of the human centrosomal kinase, Nek2, depends on an unusual leucine zipper dimerization motif. J Biol Chem 274: 16304-16310.
440. Mi J, Guo C, Brautigan DL, Larner JM (2007) Protein phosphatase-1alpha regulates centrosome splitting through Nek2. Cancer Res 67: 1082-1089.
441. Mardin BR, Lange C, Baxter JE, Hardy T, Scholz SR, et al. (2010) Components of the Hippo pathway cooperate with Nek2 kinase to regulate centrosome disjunction. Nat Cell Biol 12: 1166-1176.
442. Fry AM, Mayor T, Meraldi P, Stierhof YD, Tanaka K, et al. (1998) C-Nap1, a novel centrosomal coiled-coil protein and candidate substrate of the cell cycle-regulated protein kinase Nek2. J Cell Biol 141: 1563-1574.
443. Bahe S, Stierhof YD, Wilkinson CJ, Leiss F, Nigg EA (2005) Rootletin forms centriole-associated filaments and functions in centrosome cohesion. J Cell Biol 171: 27-33.
444. Hames RS, Crookes RE, Straatman KR, Merdes A, Hayes MJ, et al. (2005) Dynamic recruitment of Nek2 kinase to the centrosome involves microtubules, PCM-1, and localized proteasomal degradation. Mol Biol Cell 16: 1711-1724.
445. Eto M, Elliott E, Prickett TD, Brautigan DL (2002) Inhibitor-2 regulates protein phosphatase-1 complexed with NimA-related kinase to induce centrosome separation. J Biol Chem 277: 44013-44020.
446. Pugacheva EN, Golemis EA (2005) The focal adhesion scaffolding protein HEF1 regulates activation of the Aurora-A and Nek2 kinases at the centrosome. Nat Cell Biol 7: 937-946.
447. Aoki-Kinoshita KF, Kanehisa M (2007) Gene annotation and pathway mapping in KEGG. Methods Mol Biol 396: 71-91.
448. Helps NR, Luo X, Barker HM, Cohen PT (2000) NIMA-related kinase 2 (Nek2), a cell-cycle-regulated protein kinase localized to centrosomes, is complexed to protein phosphatase 1. Biochem J 349: 509-518.
449. Sardon T, Pache RA, Stein A, Molina H, Vernos I, et al. (2010) Uncovering new substrates for Aurora A kinase. EMBO Rep 11: 977-984.
450. Compton DA, Luo C (1995) Mutation of the predicted p34cdc2 phosphorylation sites in NuMA impair the assembly of the mitotic spindle and block mitosis. J Cell Sci 108 (Pt 2): 621-633.

451. Ou Y, Rattner JB (2000) A subset of centrosomal proteins are arranged in a tubular conformation that is reproduced during centrosome duplication. Cell Motil Cytoskeleton 47: 13-24.
452. Woodcock SA, Rushton HJ, Castaneda-Saucedo E, Myant K, White GR, et al. (2010) Tiam1-Rac signaling counteracts Eg5 during bipolar spindle assembly to facilitate chromosome congression. Curr Biol 20: 669-675.
453. Whitehead CM, Rattner JB (1998) Expanding the role of HsEg5 within the mitotic and post-mitotic phases of the cell cycle. J Cell Sci 111 (Pt 17): 2551-2561.
454. Sawin KE, Mitchison TJ (1995) Mutations in the kinesin-like protein Eg5 disrupting localization to the mitotic spindle. Proc Natl Acad Sci U S A 92: 4289-4293.
455. Nakagawa Y, Yamane Y, Okanoue T, Tsukita S, Tsukita S (2001) Outer dense fiber 2 is a widespread centrosome scaffold component preferentially associated with mother centrioles: its identification from isolated centrosomes. Mol Biol Cell 12: 1687-1697.
456. Ponsard C, Seltzer V, Perret E, Tournier F, Middendorp S (2007) Identification of BCAP, a new protein associated with basal bodies and centrioles. Front Biosci 12: 3683-3693.
457. Huber D, Geisler S, Monecke S, Hoyer-Fender S (2008) Molecular dissection of ODF2/Cenexin revealed a short stretch of amino acids necessary for targeting to the centrosome and the primary cilium. Eur J Cell Biol 87: 137-146.
458. He R, Wu Q, Zhou H, Huang N, Chen J, et al. (2013) Cep57 protein is required for cytokinesis by facilitating central spindle microtubule organization. J Biol Chem 288: 14384-14390.
459. Nachury MV, Maresca TJ, Salmon WC, Waterman-Storer CM, Heald R, et al. (2001) Importin beta is a mitotic target of the small GTPase Ran in spindle assembly. Cell 104: 95-106.
460. Wiese C, Wilde A, Moore MS, Adam SA, Merdes A, et al. (2001) Role of importin-beta in coupling Ran to downstream targets in microtubule assembly. Science 291: 653-656.
461. Zimmerman W, Doxsey SJ (2000) Construction of centrosomes and spindle poles by molecular motor-driven assembly of protein particles. Traffic 1: 927-934.
462. Vergnolle MA, Taylor SS (2007) Cenp-F links kinetochores to Ndel1/Nde1/Lis1/dynein microtubule motor complexes. Curr Biol 17: 1173-1179.
463. Lam C, Vergnolle MA, Thorpe L, Woodman PG, Allan VJ (2010) Functional interplay between LIS1, NDE1 and NDEL1 in dynein-dependent organelle positioning. J Cell Sci 123: 202-212.
464. Wynshaw-Boris A, Gambello MJ (2001) LIS1 and dynein motor function in neuronal migration and development. Genes Dev 15: 639-651.
465. Gruss OJ, Carazo-Salas RE, Schatz CA, Guarguaglini G, Kast J, et al. (2001) Ran induces spindle assembly by reversing the inhibitory effect of importin alpha on TPX2 activity. Cell 104: 83-93.
466. Kufer TA, Sillje HH, Korner R, Gruss OJ, Meraldi P, et al. (2002) Human TPX2 is required for targeting Aurora-A kinase to the spindle. J Cell Biol 158: 617-623.
467. Seah MK, Holt JE, Garcia-Higuera I, Moreno S, Jones KT (2012) The APC activator fizzy-related-1 (FZR1) is needed for preimplantation mouse embryo development. J Cell Sci 125: 6030-6037.
468. Wang P, Pinson X, Archambault V (2011) PP2A-twins is antagonized by greatwall and collaborates with polo for cell cycle progression and centrosome attachment to nuclei in drosophila embryos. PLoS Genet 7: e1002227.

469. Wu Q, He R, Zhou H, Yu AC, Zhang B, et al. (2012) Cep57, a NEDD1-binding pericentriolar material component, is essential for spindle pole integrity. Cell Res 22: 1390-1401.
470. Ma W, Baumann C, Viveiros MM (2010) NEDD1 is crucial for meiotic spindle stability and accurate chromosome segregation in mammalian oocytes. Dev Biol 339: 439-450.
471. Nigg EA, Stearns T (2011) The centrosome cycle: Centriole biogenesis, duplication and inherent asymmetries. Nat Cell Biol 13: 1154-1160.
472. Andersen JS, Wilkinson CJ, Mayor T, Mortensen P, Nigg EA, et al. (2003) Proteomic characterization of the human centrosome by protein correlation profiling. Nature 426: 570-574.
473. Muller H, Schmidt D, Dreher F, Herwig R, Ploubidou A, et al. (2011) Gene ontology analysis of the centrosome proteomes of Drosophila and human. Commun Integr Biol 4: 308-311.
474. Muller H, Schmidt D, Steinbrink S, Mirgorodskaya E, Lehmann V, et al. (2010) Proteomic and functional analysis of the mitotic Drosophila centrosome. EMBO J 29: 3344-3357.
475. Anthropology J-F-B-IoZa (2011) Atlas of Genetics and cytogenetics in oncology and haematology.

Protein	Pathway Map	Expression periodicities	Description	Location	Role in the Centrosome Duplication	Function
ODF2	S/G2	S G2/M	Outer dense fiber of sperm tails 2	Appendage (Distal & Subdistal)		Microtubule assembly
CEP164	M	S G2/M	Centrosomal protein 164kDa	Appendage (Distal)		Microtubule assembly
MAPRE2	M	G0	Microtubule-associated protein, RP/EB family, member 2	Appendage (Subdistal)		Microtubule assembly
MAPRE3	M	G1 E	Microtubule-associated protein, RP/EB family, member 3	Appendage (Subdistal)		Microtubule assembly
CEP170	S/G2	G1 E	Centrosomal protein 170kDa	Appendage (Subdistal)		Microtubule assembly
CNTRL	M	S G2/M	Centriolin	Appendage (Subdistal)		Microtubule assembly
MAPRE1	M	S G2/M	Microtubule-associated protein, RP/EB family, member 1	Appendage (Subdistal)		Microtubule assembly
NIN	M	S G2/M	Ninein (GSK3B interacting protein)	Appendage (Subdistal)		Microtubule assembly
KIF24	S/G2	S G2/M	Kinesin family member 24	Appendage (Subdistal)		Microtubule assembly
DCTN1	M	Consistent expressio	Dynactin 1	Centriolar satellite		Microtubule assembly
DISC1	S/G2	G1 E	Disrupted in schizophrenia 1	Centriolar satellite		Parental centriole pairing
BBS4	S/G2	G1 E	Bardet-Biedl syndrome 4	Centriolar satellite		Microtubule assembly
HOOK3	S/G2	G1 E	Hook homolog 3 (Drosophila)	Centriolar satellite		Parental centriole pairing
OFD1	S/G2	G1 E	Oral-facial-digital syndrome 1	Centriolar satellite		Parental centriole pairing
KIZ	S/G2	G1 E	Kizuna centrosomal protein	Centriolar satellite		Parental centriole pairing
CEP90	S/G2	G1 L	Centrosomal protein 90kDa	Centriolar satellite		Parental centriole pairing
PARD6A	S/G2	S G2/M	Par-6 partitioning defective 6 homolog alpha (C. elegans)	Centriolar satellite		Parental centriole pairing
CEP290	S/G2	S G2/M	Centrosomal protein 290kDa	Centriolar satellite		Microtubule assembly
CEP72	S/G2	S G2/M	Centrosomal protein 72kDa	Centriolar satellite		Parental centriole pairing
FOPNL	S/G2	S G2/M	FGFR1OP N-terminal like	Centriolar satellite		Parental centriole pairing
PCM1	S/G2	S G2/M	Pericentriolar material 1	Centriolar satellite		Microtubule assembly
MZT2B	S/G2	Consistent expressio	Mitotic spindle organizing protein 2B	Centriole		Centriole developing
CEP350	S/G2	G1 E	Centrosomal protein 350kDa	Centriole		Centriole developing
CEP55	M	S G2/M	Centrosomal protein 55kDa	Centriole		NEBD
MZT2A	S/G2	S G2/M	Mitotic spindle organizing protein 2A	Centriole		Centriole developing
CENPJ	S/G2	S G2/M	Centromere protein J	Centriole (Cartwheel structure, Proximal end)		Centriole developing
CEP135	S/G2	S G2/M	Centrosomal protein 135kDa	Centriole (Cartwheel structure, Proximal end)		Centriole developing
SASS6	S/G2	S G2/M	Spindle assembly 6 homolog (C. elegans)	Centriole (Cartwheel structure, Proximal end)		Centriole developing
CETN1	S/G2	Consistent expressio	Centrin, EF-hand protein, 1	Centriole (Distal end)	Inhibits reduplication	Microtubule assembly
ODF1	S/G2	Consistent expressio	Outer dense fiber of sperm tails 1	Centriole (Distal end)	Inhibits reduplication	Centriole developing
USP33	S/G2	Consistent expressio	Ubiquitin specific peptidase 33	Centriole (Distal end)	Inhibits reduplication	Centriole developing
POC1B	S/G2	G1 E	POC1 centriolar protein homolog B (Chlamydomonas)	Centriole (Distal end)		Centriole developing
CCNF	S/G2	S G2/M	Cyclin F	Centriole (Distal end)		Centriole developing

Appendix I. Parts list of the centrosome-regulated proteins. Pathway map: Consistent Expression: Genes that constantly express across entire cell cycle.

Protein	Pathway Map	Expression periodicities	Description	Location	Role in the Centrosome Duplication	Function
POC1A	S/G2	S G2/M	POC1 centriolar protein homolog A (Chlamydomonas)	Centriole (Distal end)		Centriole developing
POC5	S/G2	S G2/M	POC5 centriolar protein homolog (Chlamydomonas)	Centriole (Distal end)		Centriole developing
NPM1	G1	S G2/M	Nucleophosmin (nucleolar phosphoprotein B23, numatrin)	Centriole (Procentriole)		Initiates centrosome duplication
ROCK2	S/G2	G1 E	Rho-associated, coiled-coil containing protein kinase 2	Centriole (Procentriole)		Dipose NPM1 (Licence protein)
STAG3	S/G2	Consistent expressio	Stromal antigen 3	Cohesin complex		Parental centriole pairing
PP2A	S/G2	G0	Protein phosphatase 2, alpha	Cohesin complex		Parental centriole pairing
CC2D1A	S/G2	G0	Coiled-coil and C2 domain containing 1A	Cohesin complex		Parental centriole pairing
RAD21	S/G2	S G2/M	RAD21 homolog (S. pombe)	Cohesin complex		Parental centriole pairing
SGOL1	S/G2	S G2/M	Shugoshin-like 1 (S. pombe)	Cohesin complex		Parental centriole pairing
SMC1	S/G2	S G2/M	Structural maintenance of chromosomes 1A	Cohesin complex		Parental centriole pairing
SMC3	S/G2	S G2/M	Structural maintenance of chromosomes 3	Cohesin complex		Parental centriole pairing
SPAG5	S/G2	S G2/M	Sperm associated antigen 5	Cohesin complex		Parental centriole pairing
STAG2	S/G2	S G2/M	Stromal antigen 2	Cohesin complex		Parental centriole pairing
DNAH2	M	Consistent expressio	Dynein, axonemal, heavy chain 2	Cytoplasm		Microtubule assembly
TEKT1	M	Consistent expressio	Tektin 1	Cytoplasm		Microtubule assembly
CASS4	S/G2	Consistent expressio	Cas Scaffolding Protein Family Member 4	Cytoplasm	Inhibits reduplication	Paired Parental Centrosomes disjunction
CDH1	S/G2	Consistent expressio	Cadherin 1, type 1, E-cadherin (epithelial)	Cytoplasm	Inhibits reduplication	Centriole developing
PPP1CA	S/G2	Consistent expressio	Protein phosphatase 1, catalytic subunit, alpha isozyme	Cytoplasm	Inhibits reduplication	Paired Parental Centrosomes disjunction
PPP1R2	S/G2	Consistent expressio	Protein phosphatase 1, regulatory (inhibitor) subunit 2	Cytoplasm	Inhibits reduplication	Paired Parental Centrosomes disjunction
NEK6	M	G1 E	NIMA-related kinase 6	Cytoplasm		Microtubule assembly
NEK9	M	G1 E	NIMA-related kinase 9	Cytoplasm		Microtubule assembly
NDEL1	M	G1 E	NudE nuclear distribution E homolog (A. nidulans)-like 1	Cytoplasm		Microtubule assembly
SAV1	S/G2	G1 E	Salvador homolog 1 (Drosophila)	Cytoplasm		Paired Parental Centrosomes disjunction
MAPK1	M	G1 L	Mitogen-activated protein kinase 1	Cytoplasm		NEBD
CCNE1	S/G2	G1 L	Cyclin E1	Cytoplasm		Dipose NPM1 (Licence protein)
ANAPC	M	S G2/M	Anaphase promoting complex	Cytoplasm		Microtubule assembly
CDK1	M	S G2/M	Cyclin-dependent kinase 1	Cytoplasm		Microtubule assembly
FZR1	M	S G2/M	Fizzy/cell division cycle 20 related 1 (Drosophila)	Cytoplasm		Microtubule assembly
KATNB1	M	S G2/M	Katanin p80 (WD repeat containing) subunit B 1	Cytoplasm		Microtubule assembly
KPNB1	M	S G2/M	Karyopherin (importin) beta 1	Cytoplasm		Microtubule assembly
PLK1	M	S G2/M	Polo-like kinase 1	Cytoplasm		Paired Parental Centrosomes disjunction
TPX2	M	S G2/M	TPX2, microtubule-associated, homolog (Xenopus laevis)	Cytoplasm		Paired Parental Centrosomes disjunction

Protein	Pathway Map	Expression periodicities	Description	Location	Role in the Centrosome Duplication	Function
AURKA	S/G2	S G2/M	Aurora kinase A	Cytoplasm		Paired Parental Centrosomes disjunction
NEK2	S/G2	S G2/M	NIMA-related kinase 2	Cytoplasm		Paired Parental Centrosomes disjunction
NEK7	S/G2	S G2/M	NIMA-related kinase 7	Cytoplasm		Microtubule assembly
PPP1CC	S/G2	S G2/M	Protein phosphatase 1, catalytic subunit, gamma isozyme	Cytoplasm		Paired Parental Centrosomes disjunction
STK3	S/G2	S G2/M	Serine/threonine kinase 3	Cytoplasm		Paired Parental Centrosomes disjunction
CDK2	S/G2	S G2/M	Cyclin-dependent kinase 2	Cytoplasm		Dipose NPM1 (Licence protein)
CROCC	M	Consistent expressio	Ciliary rootlet coiled-coil, rootletin	Linker fibres	Inhibits reduplication	Parental centriole pairing
CEP250	M	Consistent expressio	Centrosomal protein 250kDa	Linker fibres	Inhibits reduplication	Parental centriole pairing
CEP68	M	Consistent expressio	Centrosomal protein 68kDa	Linker fibres	Inhibits reduplication	Parental centriole pairing
LRRC45	M	Consistent expressio	Leucine rich repeat containing 45	Linker fibres		Linker fibres
CTNNB1	M	G1 E	Catenin (cadherin-associated protein), beta 1, 88kDa	Linker fibres		Linker fibres
TSG101	Cytokinesis	Consistent expressio	Tumor susceptibility gene 101	Midbody		Abscission (Midbody)
PDCD6IP	Cytokinesis	G1 L	Programmed cell death 6 interacting protein	Midbody		Abscission (Midbody)
BRCA2	Cytokinesis	S G2/M	Breast cancer 2, early onset	Midbody		Abscission (Midbody)
PAFAH1B1	M	Consistent expressio	platelet-activating factor acetylhydrolase 1b, regulatory sub	PCM		Mitotic spindle assembly
NUMA1	M	G1 E	Nuclear mitotic apparatus protein 1	PCM		Mitotic spindle assembly
HAUS5	M	S G2/M	HAUS augmin-like complex, subunit 5	PCM		Mitotic spindle assembly
HAUS1	M	S G2/M	HAUS augmin-like complex, subunit 1	PCM		Mitotic spindle assembly
HAUS2	M	S G2/M	HAUS augmin-like complex, subunit 2	PCM		Mitotic spindle assembly
HAUS3	M	S G2/M	HAUS augmin-like complex, subunit 3	PCM		Mitotic spindle assembly
HAUS4	M	S G2/M	HAUS augmin-like complex, subunit 4	PCM		Mitotic spindle assembly
HAUS6	M	S G2/M	HAUS augmin-like complex, subunit 6	PCM		Mitotic spindle assembly
HAUS7	M	S G2/M	HAUS augmin-like complex, subunit 7	PCM		Mitotic spindle assembly
HAUS8	M	S G2/M	HAUS augmin-like complex, subunit 8	PCM		Mitotic spindle assembly
KIF11	M	S G2/M	Kinesin family member 11	PCM		Mitotic spindle assembly
AKAP9	S/G2	G1 L	A kinase (PRKA) anchor protein (yotiao) 9	PCM (Inner)		Maturation of centrosome
CDK5RAP2	S/G2	S G2/M	CDK5 regulatory subunit associated protein 2	PCM (Inner)		Maturation of centrosome
CEP120	S/G2	S G2/M	Centrosomal protein 120kDa	PCM (Inner)		Maturation of centrosome
PCNT	S/G2	S G2/M	Pericentrin	PCM (Inner)		Paired Parental Centrosomes disjunction
CEP63	S/G2	Consistent expressio	Centrosomal protein 63kDa	PCM (Outer)		Microtubule assembly
ASPM	S/G2	S G2/M	Asp (abnormal spindle) homolog, microcephaly associated	PCM (Outer)		Microtubule assembly

Protein	Pathway Map	Expression periodicities	Description	Location	Role in the Centrosome Duplication	Function
CEP57	S/G2	S G2/M	Centrosomal protein 57kDa	PCM (Outer)		Microtubule assembly
CNN	S/G2	S G2/M	Centrosomin	PCM (Outer)		Microtubule assembly
NEDD1	S/G2	S G2/M	Neural precursor cell expressed, developmentally down-reg	PCM (Outer)		Microtubule assembly
PLK2	G1	G1 E	Polo-like kinase 2	PCM (Parental centrosome)		Initiates centrosome duplication
CEP192	G1	S G2/M	Centrosomal protein 192kDa	PCM (Parental centrosome)		Initiates centrosome duplication
PLK4	G1	S G2/M	Polo-like kinase 4	PCM (Parental centrosome)		Initiates centrosome duplication
STIL	G1	S G2/M	SCL/TAL1 interrupting locus	PCM (Parental centrosome)		Initiates centrosome duplication
TUBGCP2	S/G2	Consistent expressic	Tubulin, gamma complex associated protein 2	TuRCs (Microtubules)		Microtubule assembly
TUBGCP6	S/G2	Consistent expressic	Tubulin, gamma complex associated protein 6	TuRCs (Microtubules)		Microtubule assembly
TUBA1B	S/G2	S G2/M	Tubulin, alpha 1b	TuRCs (Microtubules)		Microtubule assembly
TUBB3	S/G2	S G2/M	Tubulin, beta 3 class III	TuRCs (Microtubules)		Microtubule assembly
TUBD1	S/G2	S G2/M	Tubulin, delta 1	TuRCs (Microtubules)		Microtubule assembly
TUBG1	S/G2	S G2/M	Tubulin, gamma 1	TuRCs (Microtubules)		Microtubule assembly
TUBGCP3	S/G2	S G2/M	Tubulin, gamma complex associated protein 3	TuRCs (Microtubules)		Microtubule assembly
TUBGCP4	S/G2	S G2/M	Tubulin, gamma complex associated protein 4	TuRCs (Microtubules)		Microtubule assembly
TUBGCP5	S/G2	S G2/M	Tubulin, gamma complex associated protein 5	TuRCs (Microtubules)		Microtubule assembly
TUBE1	S/G2	G0	Tubulin, epsilon 1	TuRCs (Microtubules) & Subdistal appendage		Microtubule assembly

Appendix II. R source code used for between Array normalisation and annotation.

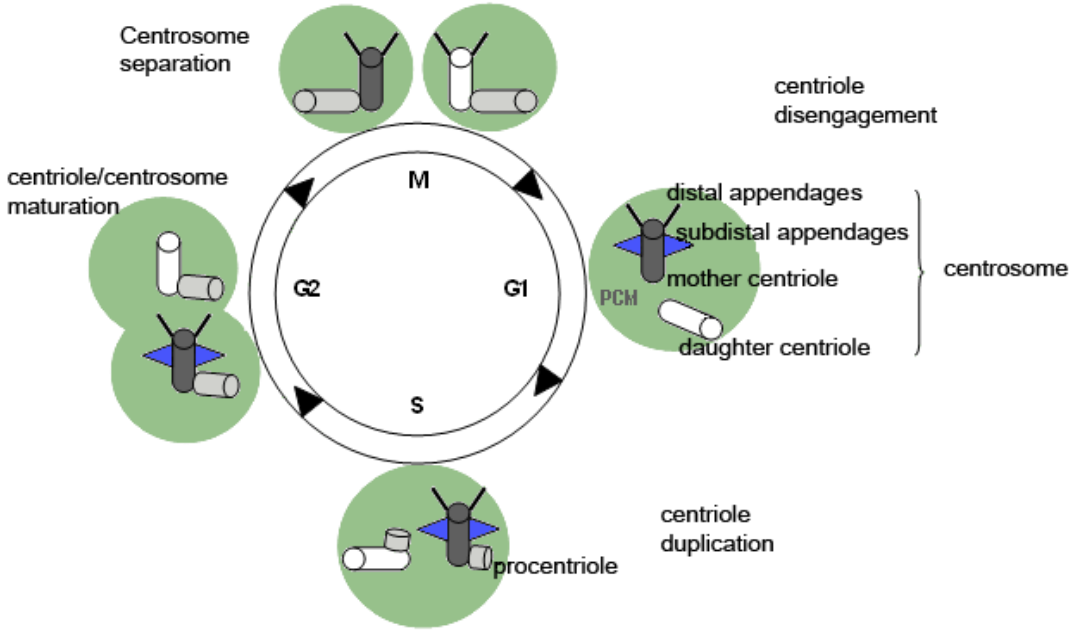
R command lines	
After: Appendix	Source code for R x64 3.0.1 with oligo version 1.24.2
# Using oligo packages for RMA normalisation.	library(oligo) cels <- list.celfiles() celfiles <- read.celfiles(cels) celfiles.norm <- rma(celfiles) eset <- as.data.frame(exprs(celfiles.norm)) eset[,c(1:40)] <- 2 ** eset[,c(1:40)] probe <- rownames(eset)
# Annotation of transcripts using packages from Affymetrix hugene11 annotation data.	eset$probe <-probe library(hugene11sttranscriptcluster.db)
# source ("https://bioconductor.org/biocLite.R")	symbols <- mget(probe, hugene11sttranscriptclusterSYMBOL, ifnotfound=NA)
# biocLite ("hugene11sttranscriptcluster.db")	entrezids <- mget(probe, hugene11sttranscriptclusterENTREZID, ifnotfound=NA) descriptions <- mget(probe, hugene11sttranscriptclusterGENENAME, ifnotfound=NA)
# Reorganise terms for transcripts annotation using R function 'sapply' and 'paste'.	eset$Symbol <- unlist(symbols) eset$EntrezID <- unlist(entrezids) eset$Description <- unlist(descriptions) library(hugene11sttranscriptcluster.db) ENSEMBL <- sapply(contents(hugene11sttranscriptclusterENSEMBL), paste, collapse=", ") ENSEMBL<- as.data.frame(ENSEMBL) ENSEMBL$probe <- row.names(ENSEMBL) ENSEMBL2PROBE<- as.data.frame(hugene11sttranscriptclusterENSEMBL2PROBE) ENZYME <- sapply(contents(hugene11sttranscriptclusterENZYME), paste, collapse=", ") ENZYME<-as.data.frame(ENZYME) ENZYME$probe <-row.names(ENZYME) ENZYME2PROBE <- as.data.frame(hugene11sttranscriptclusterENZYME2PROBE) GO <-sapply(contents(hugene11sttranscriptclusterGO), paste, collapse=", ") GO <- as.data.frame(GO) GO$probe <-row.names(GO) GO2ALLPROBES<- as.data.frame(hugene11sttranscriptclusterGO2ALLPROBES) colnames(GO2ALLPROBES)[1]<- "probe" GO2ALLPROBES_2<-GO2ALLPROBES[,1:2] GO2ALLPROBES_3<-as.data.frame(GO2ALLPROBES_2) GO2ALLPROBES_4<- sapply(data.frame(GO2ALLPROBES_2), paste, collapse=", ") GO2ALLPROBES <- sapply(contents(hugene11sttranscriptclusterGO2ALLPROBES), paste, collapse=", ") GO2ALLPROBES<- sapply(contents(hugene11sttranscriptclusterGO2ALLPROBES),

```
                                    paste, collapse=", ")
                                    GO2PROBE <-
                                    sapply(contents(hugene11sttranscriptclusterGO2PROBE), paste,
                                    collapse=", ")
                                    MAP <-sapply(contents(hugene11sttranscriptclusterMAP),
                                    paste, collapse=", ")
                                    MAP<-as.data.frame(MAP)
# Annotate transcripts using R      MAP$probe<-rownames(MAP)
  function 'merge'.                 OMIM <-sapply(contents(hugene11sttranscriptclusterOMIM),
                                    paste, collapse=", ")
                                    OMIM<-as.data.frame(OMIM)
                                    OMIM$probe<-rownames(OMIM)
                                    PATH <-sapply(contents(hugene11sttranscriptclusterPATH),
                                    paste, collapse=", ")
                                    PATH <-as.data.frame(PATH)
                                    PATH$probe<-rownames(PATH)
                                    PATH2PROBE <-
                                    sapply(contents(hugene11sttranscriptclusterPATH2PROBE),
                                    paste, collapse=", ")
                                    PFAM <-sapply(contents(hugene11sttranscriptclusterPFAM),
                                    paste, collapse=", ")
                                    PFAM<-as.data.frame(PFAM)
                                    PFAM$probe<-rownames(PFAM)
# Save files as '.csv' format for   PMID <-sapply(contents(hugene11sttranscriptclusterPMID),
  further analyses.                 paste, collapse=", ")
                                    PMID<-as.data.frame(PMID)
                                    PMID$probe<-rownames(PMID)
                                    PMID2PROBE <-
                                    sapply(contents(hugene11sttranscriptclusterPMID2PROBE),
                                    paste, collapse=", ")
                                    PROSITE <-
                                    sapply(contents(hugene11sttranscriptclusterPROSITE), paste,
                                    collapse=", ")
                                    PROSITE <-as.data.frame(PROSITE)
                                    PROSITE $probe<-rownames(PROSITE)
                                    UNIGENE <-
                                    sapply(contents(hugene11sttranscriptclusterUNIGENE), paste,
                                    collapse=", ")
                                    UNIGENE <-as.data.frame(UNIGENE)
                                    UNIGENE $probe<-rownames(UNIGENE)
                                    UNIPROT <-
                                    sapply(contents(hugene11sttranscriptclusterUNIPROT), paste,
                                    collapse=", ")
                                    UNIPROT <-as.data.frame(UNIPROT)
                                    UNIPROT $probe<-rownames(UNIPROT )

                                    merge(eset,ENSEMBL,by.x="probe",by.y="probe",all.x=TRUE)
                                    ->eset
                                    merge(eset,ENSEMBL2PROBE,by.x="probe",by.y="probe",all.
                                    x=TRUE)->eset
                                    merge(eset,ENZYME,by.x="probe",by.y="probe",all.x=TRUE)-
                                    >eset
                                    merge(eset,ENZYME2PROBE,by.x="probe",by.y="probe",all.x
                                    =TRUE)->eset
                                    merge(eset,GO,by.x="probe",by.y="probe",all.x=TRUE)->eset
                                    merge(eset,GO2ALLPROBES,by.x="probe",by.y="probe",all.x=
                                    TRUE)->eset
```

```
merge(eset,GO2PROBE,by.x="probe",by.y="probe",all.x=TRU
E)->eset
merge(eset,MAP,by.x="probe",by.y="probe",all.x=TRUE)->eset
merge(eset,OMIM,by.x="probe",by.y="probe",all.x=TRUE)-
>eset
merge(eset,PATH,by.x="probe",by.y="probe",all.x=TRUE)-
>eset
merge(eset,PATH2PROBE,by.x="probe",by.y="probe",all.x=TR
UE)->eset
merge(eset,PFAM,by.x="probe",by.y="probe",all.x=TRUE)-
>eset
merge(eset,PMID,by.x="probe",by.y="probe",all.x=TRUE)-
>eset
merge(eset,PMID2PROBE,by.x="probe",by.y="probe",all.x=TR
UE)->eset
merge(eset,PROSITE,by.x="probe",by.y="probe",all.x=TRUE)-
>eset
merge(eset,UNIGENE,by.x="probe",by.y="probe",all.x=TRUE)
->eset
merge(eset,UNIPROT,by.x="probe",by.y="probe",all.x=TRUE)-
>eset

write.csv(eset,
"Labelled_RMA_normalised_33297_transcripts.csv")
```



Appendix III Centrosome duplication cycle. This figure is courtesy of Atlas of Genetics and Cytogenetics in Oncology and Haematology [475].